OPERATION UNDER UNCERTAINTY IN ELECTRIC GRID: A MULTIPARAMETRIC PROGRAMMING APPROACH

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yuting Ji

January 2017

OPERATION UNDER UNCERTAINTY IN ELECTRIC GRID: A MULTIPARAMETRIC PROGRAMMING APPROACH

Yuting Ji, Ph.D.

Cornell University 2017

Uncertainty is a major factor in power system operations. In recent years, with the emergence of the smart grid, uncertainty level has been further elevated in both the generation and demand side of power systems. Increasing uncertainty exposes the electric grid to potential safety issues and economic loss, thus posing significant challenges to the grid operations.

Traditionally, power system operations use certainty equivalent approach to deal with uncertainty, *i.e.*, replacing random variables by their expected values. With this simplification, the original stochastic optimization is reduced to a deterministic problem. However, the certainty equivalent method is inadequate for the modern electric grid with deep penetration of distributed energy resources. Due to increasing uncertainty, operations and decision makings need to incorporate system dynamics over a broad range of temporal and spatial horizons. To this end, this thesis provides a new paradigm for operation under uncertainty and computationally efficient algorithms based on multiparametric programming theory. Under this new paradigm, uncertainty is characterized by conditional distributions and decisions are made by incorporating such probabilistic descriptions. To illustrate the new paradigm, we consider two specific problems. For characterization of system uncertainty, we develop a formal methodology for probabilistic forecasting of real-time operations and locational

marginal prices. Conditioning on the current system state, we provide a full distribution of future operations and prices. For operational decision making, we propose an optimal stochastic approach to interchange scheduling in multi-area systems. By incorporating the conditional distribution of load and generation, the optimal interchange is obtained through an iterative process.

BIOGRAPHICAL SKETCH

Yuting Ji received the bachelor degree in computer science from Tsinghua University, Beijing, China, in 2011. She is working toward the Ph.D. degree in electrical and computer engineering, Cornell University, Ithaca, NY, USA.

During the summer in 2012, she was a student intern with IBM Thomas J. Watson Research Center, Hawthorne, NY, USA. She was a recipient of Jacobs Fellowship of Cornell University in 2011 and a finalist for the Best Conference Paper in 2016 IEEE Power & Energy System General Meeting. Her research interests include optimization, learning, statistical inference, smart grid and power system.

For my family.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the many people I am indebted to, and I would like to thank them here.

First and foremost, I would like to thank my advisor, Professor Lang Tong. He has been supportive since the first day I started as a graduate student at Cornell. He has supported me financially, academically and emotionally through the road to this thesis, and trained me from a computer science student who probably only knows Kirchhoff's laws to an expert in a tiny subfield of power system. I would like to thank him for his patience with my inexperience, his sharp critique on my ideas, his encouragement at my times of depression, and his devotion to mentoring me. I hope that I could be as professional, enthusiastic, energetic and humorous as him and to become a wise adviser like him in the future.

I am extremely grateful to have Professors Robert Thomas, Tim Mount, and Eilyan Bitar, for serving on my committee and providing helpful comments on my research during my candidacy exam, defense, CERTS reviews and E3RG presentations.

Dozens of people have helped and taught me immensely during my graduate study. I would like to express my gratitude to Ting He for teaching me how to write papers and for her guidance during my internship at IBM, amazing colleagues, Jinsub Kim, Shiyao Chen, Liyan Jia, Zhe Yu, Daniel Munoz-Alvarez, Mukadder Sevi Baltaoglu, and Ye Guo, for generous help and many interesting research discussions, and Tongxin Zheng from ISO New England for fruitful collaboration and valuable comments. Among my long list of remarkable per-

sonal friends, I sincerely thank Nan Li for all the encouragement and help in my last year of graduate school.

Finally, I would like to thank my family for their unconditional love and support. They have cherished with me every great moment and supported me whenever I needed it. I cannot express my gratefulness to my family in words.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The electric grids are continuously exposed to numerous sources of uncertainties which threaten the grid reliability and safety. In traditional power systems, the uncertainty mainly exists in future energy consumption and the electric component outages, such as unit outage, transmission line breakdown, and breaker faults. In recent years, with the emergence of the smart grid, uncertainty level has been further elevated in both the generation and demand side of power systems. Variable renewable electricity generation capacity in the United States has increased considerably. Wind generation capacity, for example, has increased from 2.6 GW in 2000 to 40 GW in 2010 [39]. This trend will continue in part because of federal incentives and renewable portfolio standards mandated in many states. New York, for example, approves the most ambitious clean energy standard mandating 50% of power from renewables by 2030 [6].

Increasing renewable energy penetration poses significant challenges to the grid operations. The difficulty associated with integrating variable sources of electricity stems from the fact that the output of such generation is highly *uncertain and uncontrollable*. Since the balance between electricity supply and demand must be maintained at all times, renewable energy fluctuation forces the operator to adjust conventional generation to compensate. Take the example of solar panels. The famous "duck curve", shown in Figure 1.1, presents sharp ramps and frequent fluctuations in net load caused by solar power generation. On a daily basis, fluctuations caused by sunrise and sunset requires subtle resource schedule. On an hourly basis, variability caused by clouds can make it difficult

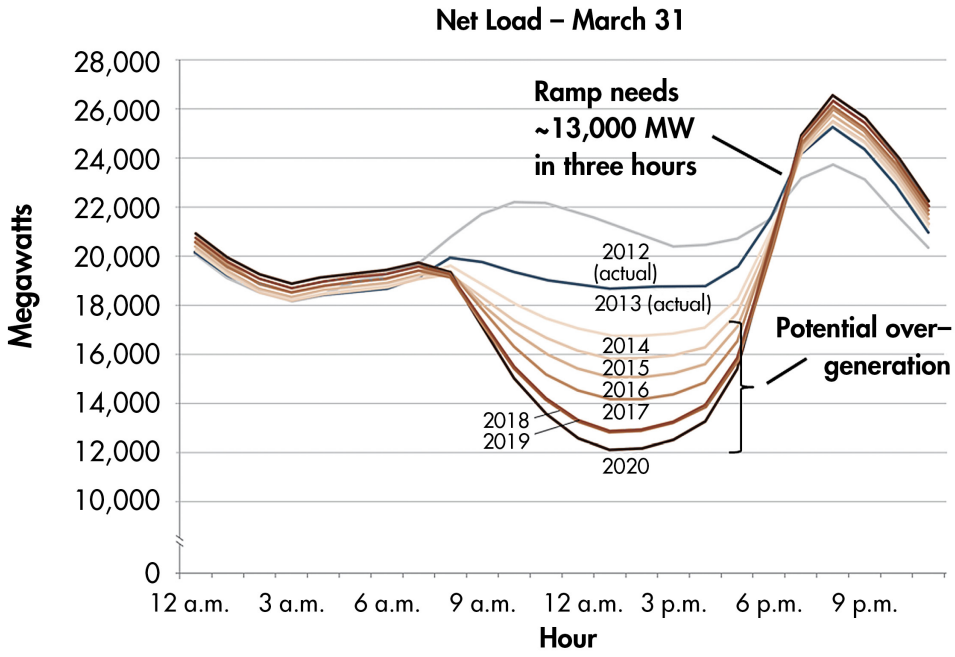


Figure 1.1: CAISO net load curve on March 31, 2013.

for the grid operator to predict how much electricity is required to compensate sudden solar generation shortfalls or excesses. It should be noted that the duck curve only shows system dynamics at the aggregate level, the actual spatial and temporal complexity is much higher than what we observe from Figure 1.1.

One of the main implications of duck curve is that the traditional approach, *i.e.*, certainty equivalent method, is inadequate to deal with the current level of uncertainty in the modern grid. A new paradigm is needed to incorporate the actual distributed dynamics into system operations. Under this new paradigm, stochastic models have to be used to describe the uncertainty in a *probabilistic* way. New solution approaches to operations and decision makings need to be developed with the consideration of probabilistic description of system uncertainty.

In this thesis, we present a new paradigm with the characterization of system uncertainty and model-specific solution approaches to system operations. A multiparametric programming approach is taken to solve problems related to *optimization under uncertainty* in power systems. A broad set of interrelated decisions in system operations can be formulated as this type of optimization problems. In particular, two specific problems are considered: (1) *probabilistic forecasting* of *real-time* electricity price and operations and (2) *stochastic* interchange scheduling in *multi-area* power systems. The former provides a probabilistic description of electricity prices and the latter develops solution approach by incorporating the conditional distribution of the random prices. Both problems share the same ingredient — optimization of power flow with the presence of system and operation uncertainty. For each problem, we identify parameters, formulate multiparametric programs, and solve them in a computationally efficient way using optimization techniques and learning theory.

1.1 Multiparametric Programming

Multiparametric programming is a technique that, in an optimization framework with a number of bounded parameters affecting the solution, obtains the exact mapping of the optimal solution from the parameter space. The optimal solution mapping consists of

1. the objective function and the decision variables (of both primal and dual problems) as functions of the parameters, and

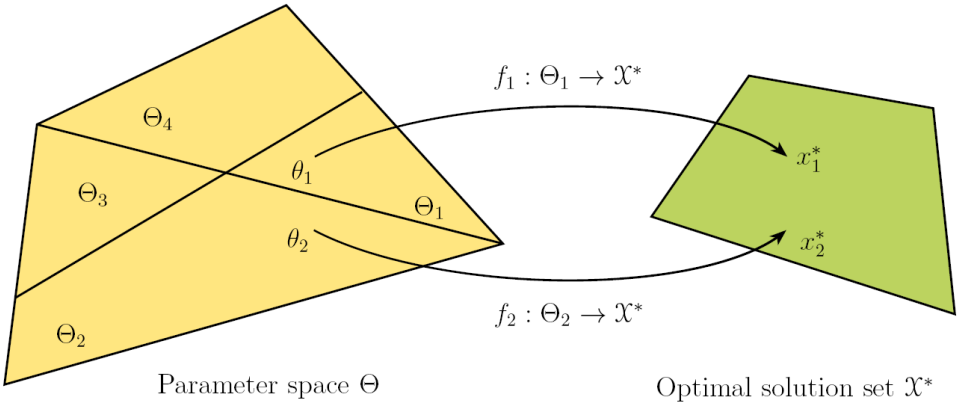


Figure 1.2: Geometric illustration for affine mappings.

2. the space of parameters (known as *critical regions*) where these functions are valid.

The optimization can then be replaced by its optimal solution mapping and the optimal solution for a given value of the parameters can be computed efficiently by performing simple function evaluations, *without solving the original program*. This thesis mainly focuses on the multiparametric linear programming for linear programs and multiparametric quadratic programming for strictly convex quadratic programs. A key result of the multiparametric linear programs and multiparametric quadratic programs is that the optimal solution is an affine function of the parameter within a critical region as shown in Figure 1.2.

For problems of optimization under uncertainty, a parametric framework is proposed by treating system random variables as parameters where suitable stochastic models are used to describe system uncertainty. By exploiting explicitly the optimal solution structure, the problem of collecting statistics on the space of *continuous probability distributions* of random parameters is reduced to

that on the space of *finite discrete probability distributions* on a set of critical regions. Since each critical region is associated with a unique affine function that maps the parameter to the optimal solution, there is no need to repetitively solve the original optimization problem for each realization of the random variable.

1.2 Probabilistic Forecasting of Real-Time Locational Marginal Price and Network Congestion

We consider the problem of short-term forecasting of locational marginal price (LMP) and transmission congestion in a deregulated electricity market. Accurate forecasts are desirable for both market participants and system operators.

For market participants, forecast of real-time prices is valuable in risk management, bidding strategy and demand side participation. The forecast price signal allows market participants make adjustments in advance to ensure economic transactions. In particular, if the forecast price at a future time is high due to, for example, the anticipated high level of demand, larger consumers or demand response participants could reduce their future consumption. Similarly, a high future price forecasts also motivate generations from suppliers.

For system operators, forecast of transmission congestion is important in congestion management, system planning and operation. Congestion management involves precautionary as well as remedial action on system operator's side. The forecast of congestion patterns indicates possible future system con-

ditions in which system operators can take reasonable actions accordingly. European transmission system operators, for instance, use Intraday Congestion Forecast (IDCF) to improve real-time security assessment [1][2].

On the other hand, LMP forecasts can also be helpful in congestion alleviation. A price spike usually happens with the presence of congestion. System operators can use its forecast as a signal to motivate consumers to adjust their usage, which in turn benefits the grid reliability. Currently, there are some system operators providing real-time price forecasts. The Electric Reliability Council of Texas (ERCOT) [22] offers a 1 hour ahead real-time LMP forecast, updated every 5 minutes. The Alberta Electric System Operator (AESO) [9] provides two near-term price forecasts with prediction horizons of 2 hours ahead and 6 hours ahead.

An accurate forecast of congestion and price can also contribute to the efficiency and competitiveness of the electricity market. To improve market efficiency, demand response programs are being used by system operators as resource options for balancing supply and demand. For real-time price-based demand response, the real-time LMP forecast is the foundation. Such forecasts can also make the market more competitive by affecting real-time bidding strategies. For example, in order to enter the market, suppliers tend to reduce their cost during those periods when future prices are relatively high. This incentive leads to a reduction of overall cost during peak times.

Since real-time LMP is stochastic in nature and may be highly volatile, its forecast is most useful to decision makers if the forecast not given as a spe-

cific value (*i.e.* the point forecast) but the probability distribution of all possible prices. This motivates the need of developing probabilistic forecasting methodologies for real-time LMP. Similarly, a probabilistic forecast of congestion pattern provides richer information for system operators.

There are significant technical challenges of probabilistic forecast of real-time congestion and LMPs. First, there needs to be an reasonably accurate model for the real-time dispatch and LMP from which probabilistic forecast can be derived. Second, there needs to be a way to incorporate real-time network operating conditions and uncertainties. Finally, the forecast algorithm needs to be reasonably simple with as much computation performed off-line as possible. These challenges are daunting if the forecaster is merely a market participant without access to network operating conditions and network parameters. On the other hand, if it is the system operator performing the forecast, as in the case of ERCOT or AESO, the barrier to efficient and accurate forecast is significantly reduced.

The real-time LMP and congestion forecasting problem from an operator perspective is considered in Chapter 3. We focus on probabilistic forecasting that, at time t, provides the conditional probability distribution at time $t + T$ of the LMP vector and associated congestion, given the system state at time t.

The key idea behind the proposed approach is the use of multiparametric programming that partitions the uncertainty space into critical regions with each region attached to a unique LMP and congestion pattern. Thus, the problem of probabilistic forecasting reduces to computing the probabilities of ran-

dom parameters falling in the set of critical regions. When loads or generations (treated as negative loads) are random, their forecasts are incorporated to generate probabilistic LMP and congestion forecasts. Computationally, the proposed method shifts a majority of the computation offline, which significantly reduces online computational cost.

An alternative algorithm that dynamically generates critical regions is also proposed. Because load and stochastic generation processes are physical processes, they are bounded and tend to concentrate around their mean trajectories. Their realizations thus only fall in a few critical regions instead of all over the entire space. By generating the critical regions that contain such realizations, the computational cost is reduced by several orders of magnitude comparing with standard Monte Carlo techniques.

1.3 Multi-Area Interchange Scheduling

Since the restructuring of the electric power industry, independent system operators (ISOs) and regional transmission organizations (RTOs) have been created to operate power grids defined by certain geographical boundaries as shown in Figure 1.3. Within each control region, a centralized market is administered by the operator who collects supply offers and demand bids, determines market clearing prices, and settles auctions for producers and consumers. On the boundary, trades between two regions are jointly determined by both operators to ensure their physical deliveries. Such inter-regional trades enable market participants to buy electricity from one region and sell it to the other, and

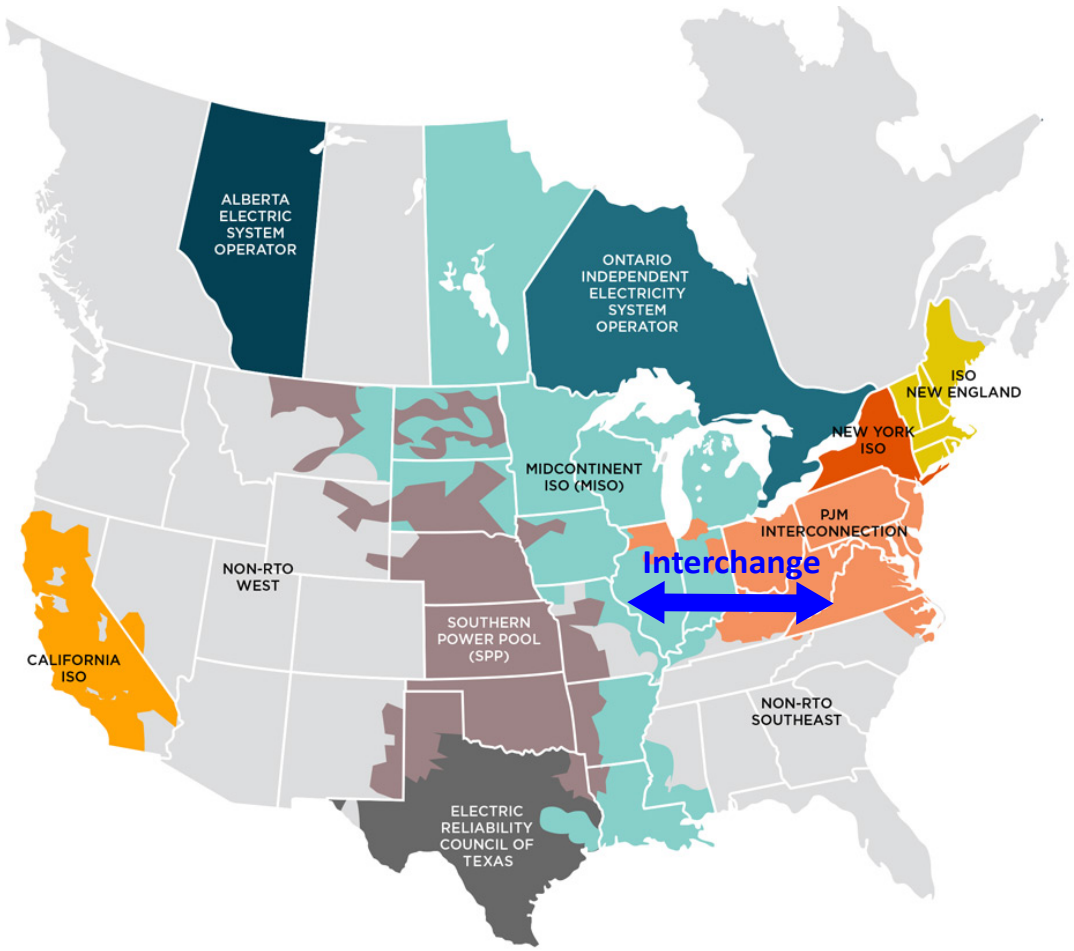


Figure 1.3: ISOs, RTOs and interchange[1].

allow low-cost external resources to compete with internal resources to serve consumers.

If there were no trading barrier across two neighboring areas, there would have been a seamless market as a whole, approximating a single larger market. In practice, however, the so-called seams between neighboring operating regions do exist, which is manifested by the counter-intuitive flows that export

[1]Source: Sustainable FERC Project — ISO RTO Operating Regions.

power from high cost to low cost regions. Another symptom of seams is the under-utilization of the interface transfer capacity when more power could have been exported from a low cost region to a high cost one.

The current industrial approach is a market-based solution that sets the net interchange[2] based on bids and offers submitted by market participants under a set of complicated rules and procedures. In [47], the analysis of the seams issues between New York ISO and ISO New England shows that the economic loss due to seams for the New York and New England customers is estimated at the level of $784 million from 2006 through 2010.

The authors of [47] point out that the latency between the scheduling of interchange and the actual power delivery is a major cause of inefficiency. Typically, the net interchange is set by a market clearing process in which external transaction offers and bids are submitted far ahead of the time of power transfer. Consequently, the interchange determined in advance may not reflect the actual system conditions. This situation is likely to be exacerbated with the greater integration of renewables.

A second factor that causes inefficiency is the lack of coordination between neighboring ISOs in their respective clearing processes of bids and offers from external market participants. This suggests that considerable gain in performance may be realized by a process that schedules the interchange based on the minimization of overall system cost.

[2]By the net interchange we mean the total power transferred from one region to another over the boundary.

An optimal interchange scheduling *in the presence of system and operation uncertainty* is first proposed for two-area single interface systems in Chapter 4. In particular, a two-stage stochastic optimization is formulated to minimize the expected overall system cost. The proposed optimization framework takes into account random fluctuations of load and renewable generation in the system. Because directly solving the stochastic optimization is intractable (with continuous uncertainty), the stochastic problem is transformed into an equivalent but deterministic optimization. This transformation allows us to generalize the deterministic solution (the current industrial practice) by intersecting the expected demand and supply functions of interchange, therefore avoiding repetitive computation and iterative information exchange between operators.

The generalization of the single interface scheduling problem is then made to the case involving a network of operating areas, each having multiple interfaces with its neighbors. A new scheduling technique based on the classical idea of coordinate descent method is proposed in [34]. The main idea is to iterate over all interfaces, one at a time, minimizing the overall system cost under uncertainty. In each iteration, a single interface optimization is solved by projecting the current solution to a particular coordinate representing a particular interface flow.

Two types of interchange scheduling are considered: (i) the synchronous scheduling and (ii) the asynchronous scheduling. The former requires all areas operated under the same scheduling clock whereas the latter allows every pair of operating areas setting their interfaces independently of others. The optimality and convergence are guaranteed under nominal assumptions for both

synchronous and asynchronous algorithms. To the best knowledge, there is no existing results on this problem in the open literature.

1.4 Summary of Contributions

In this thesis, we present a new paradigm for power operations under uncertainty. A stochastic approach to various real-time operation problems is proposed based on the theory of multiparametric programming. For each problem, we identify parameters, formulate multiparametric programs, and solve them in a computationally efficient way using additional optimization techniques and learning theory.

The main contributions of this thesis are summarized below.

1. A new paradigm for system operations is presented with two key components: characterization of uncertainty and model-specific solution approach by incorporating probabilistic descriptions of randomness.
2. A fundamental multiparametric programming framework is proposed for problems of optimization under uncertainty in the real-time operations.
3. A first attempt is made to develop a formal electricity price forecasting methodology used by a system operator with the goal of facilitating actionable information for the system operator and market participants. This probabilistic forecasting technique has the following desirable features.
    - Informative: provide joint and marginal distributions of nodal price.

- Flexible: can be applied to various models and incorporate different types of randomness.
- Efficient: satisfy the computational need on minute basis.
- Scalable: demonstrate the performance on a real-world-size power network.

4. An *optimal stochastic* interchange scheduling algorithm is proposed for multi-area multi-interface systems based on the current industrial practice.

1.5 Thesis Outline

The thesis is structured as follows.

The first part of the thesis is an introduction to multiparametric programming. In this thesis, parametric programming is the main technique used to characterize the impact of system uncertainty on the real-time operation in the electric grid. Specifically, a parametric framework is proposed for economic dispatch in which the system random variables are formulated as parameters. The optimal solution structures are explicitly exploited as algebraic functions of random variables. This parametric framework and optimal solution structures constitute the basic tools for operations under uncertainty in the electric grid.

In the second part of the thesis we focus on probabilistic forecasting of real-time LMP and network congestion. A new probabilistic forecasting technique is

proposed based on a multiparametric programming formulation that partitions the uncertainty parameter space into critical regions from which the conditional probability distribution of the real-time LMP/congestion is obtained. The proposed method incorporates load/generation forecast, time varying operation constraints, and contingency models. By shifting the computation associated with multiparametric programs offline, the online computational cost is significantly reduced. An online forecasting technique by generating critical regions dynamically is also proposed, which results in several orders of magnitude improvement in the computational cost over standard Monte Carlo methods.

In the third part of the thesis we focus on the multi-area interchange scheduling under uncertainty. An interchange scheduling technique based on a two-stage stochastic minimization of expected operating cost is proposed. The multiparametric framework is used in the second stage by formulating interchange and random load and generation as parameters. To solve the stochastic optimization for optimal interchange, an equivalent problem that maximizes the expected social welfare is formulated for each separate scheduling interface. The proposed technique leverages the operator's capability of forecasting locational marginal prices and obtains the optimal interchange schedule without iterations among operators.

For the scheduling problem in a multi-area multi-interface system, an iterative algorithm is also proposed based on the coordinate descent method. In particular, the proposed algorithm iteratively optimizes the interface flows using a multidimensional demand and supply functions. Optimality and convergence are guaranteed for both synchronous and asynchronous scheduling under nom-

inal assumptions.

We conclude this thesis in the last chapter.

CHAPTER 2

MULTIPARAMETRIC PROGRAMMING

Parametric programming is a type of mathematical optimization, where the optimization problem is solved as a function of one or multiple parameters. In contrary to sensitivity analysis, which characterizes the change of the solution with respect to small perturbations of the parameters, parametric programming systematical analyzes the effect of uncertainty and variability in mathematical programming problems.

In general, parameters may reside in the objective function or (and) the constraints of an optimization. In this thesis, we focus on a special type of parametric programs — the right-hand side multiparametric programs with linear constraints.

In this chapter, we introduce the key concepts and main results of multiparametric programming which provide the theoretical foundations for the proposed parametric framework. The definitions and theoretical results follow chapter "Multiparametric Programming: a Geometric Approach" in [16].

2.1 Problem Formulation and Critical Region

Consider a general right-hand side multiparametric program as follows:

$$\min_x z(x) \text{ subject to } Ax \leq b + E\theta \quad (y) \tag{2.1}$$

where x is the decision vector, θ the parameter vector, $z(\cdot)$ the cost function, y the Lagrangian multiplier vector, and A, E, b are coefficient matrix/vector with compatible dimensions.

We denote Θ the region of parameters such that the mathematical program (2.1) is feasible. For any given $\theta \in \Theta$, let $z^*(\theta)$, $\mathcal{X}^*(\theta)$ and $\mathcal{Y}^*(\theta)$ be the optimal objective value, the set of optimal primal solutions and the set of optimal dual solutions in problem (2.1) for θ.

In general, there may exist multiple optimal solutions to (2.1). In this thesis, we focus on the special case where the optimization (2.1) has a unique optimal primal solution and a unique optimal dual solution for each parameter value θ. The uniqueness can be guaranteed by the assumption that problem (2.1) is neither primal degenerate nor dual degenerate for all $\theta \in \Theta$. The definition of primal and dual degeneracy is given below.

Definition 1. *([16], p. 25, 32) For any given $\theta \in \Theta$, the mathematical program (2.1) is said to be primal degenerate if there exists a $x^*(\theta) \in \mathcal{X}^*(\theta)$ such that the number of active constraints at the optimizer is greater than the dimension of parameter θ.*

Definition 2. *([16], p. 26, 32) For any given $\theta \in \Theta$, the mathematical program (2.1) is said to be dual degenerate if its dual problem is primal degenerate.*

The multiparametric programming analysis builds on the concept of critical region. A critical region is a set of the parameters space where the local conditions for optimality of a multiparametric program remain unchanged.

Definition 3. *([16], p. 112) Let $\mathcal{J}$ denotes the set of constraint indices in (2.1). For any subset $\mathcal{I} \subseteq \mathcal{J}$, let $A_\mathcal{I}$ and $E_\mathcal{I}$ be the corresponding submatrices of A and E, respectively,*

consisting of rows indexed by $\mathcal{I}$. An optimal partition of the index set $\mathcal{J}$ associated with parameter $\theta \in \Theta$ is the partition $(\mathcal{I}(\theta), \mathcal{I}^c(\theta))$ where

$$\mathcal{I}(\theta) \triangleq \{i \in \mathcal{J} | A_i x^*(\theta) = b + E_i\theta, \text{ for all } x^*(\theta) \in \mathcal{X}^*(\theta)\},$$

$$\mathcal{I}^c(\theta) \triangleq \{i \in \mathcal{J} | A_i x^*(\theta) < b + E_i\theta, \text{ for some } x^*(\theta) \in \mathcal{X}^*(\theta)\}.$$

Definition 4. *([16], p. 112) For a given $\theta_0 \in \Theta$, let $(\mathcal{I}_0, \mathcal{I}_0^c) \triangleq (\mathcal{I}(\theta_0), \mathcal{I}^c(\theta_0))$. The critical region related to the index set $\mathcal{I}_0$ is defined as*

$$\Theta_{\mathcal{I}_0} \triangleq \{\theta \in \Theta | \mathcal{I}(\theta) = \mathcal{I}_0\},$$

which is the set of all parameters $\theta \in \Theta$ with the same active constraint set $\mathcal{I}_0$ at the optimum of problem (2.1).

By Definition 3, the optimal partition specifies two sets of constraints: one set is a combination of active constraints at the optimum of problem (2.1) and the other inactive. By Definition 4, a critical region is essentially a subset of the parameter space Θ on which a certain set of constraints is active at the optimum of problem (2.1).

The multiparametric programming determines the optimal solution structure of problem (2.1) consisting of

1. the feasible parameter space Θ and its critical region partition $\{\Theta_i\}$,
2. the optimal objective function $z^*(\theta)$, the optimal primal solution[1] $x^*(\theta)$, and the optimal dual solution $y^*(\theta)$ as functions of the parameter $\theta \in \Theta$.

In this thesis, we concentrate on multiparametric linear programs (MPLPs) and multiparametric quadratic programs (MPQPs).

[1] If there are multiple optimizers, only one of them $x^*(\theta) \in \mathcal{X}^*(\theta)$ is determined.

2.2 Multiparametric Linear Programming

Consider the special case of the multiparametric program (2.1) where the objective is linear

$$\min_x c^\intercal x \text{ subject to } Ax \leq b + E\theta. \quad (y) \tag{2.2}$$

Theorem 1 ([16], p. 118)**.** *Given a parameter θ_0, let $(\mathcal{I}_0, \mathcal{I}_0^c)$ be the optimal partition of the index set $\mathcal{I}$ of MPLP (2.2) associated with θ_0. Let $A_{\mathcal{I}_0}, E_{\mathcal{I}_0}$ and $b_{\mathcal{I}_0}$ be, respectively, the submatrices of A, E and subvector of b corresponding to the index set $\mathcal{I}_0$. Let $A_{\mathcal{I}_0^c}, E_{\mathcal{I}_0^c}$ and $b_{\mathcal{I}_0^c}$ be similarly defined for the index set $\mathcal{I}_0^c$. Assume that MPLP (2.2) is neither primal nor dual degenerate for all $\theta \in \Theta$. Denote the critical region that contains θ_0 by $\Theta_{\mathcal{I}_0}$.*

1. *The critical region $\Theta_{\mathcal{I}_0}$ is given by*

$$\Theta_{\mathcal{I}_0} = \left\{\theta \middle| A_{\mathcal{I}_0^c} A_{\mathcal{I}_0}^{-1}(b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta) < b_{\mathcal{I}_0^c} + E_{\mathcal{I}_0^c}\theta\right\}. \tag{2.3}$$

2. *The optimal primal solution function $x^*(\theta)$ for $\theta \in \Theta_{\mathcal{I}_0}$ is given by*

$$x^*(\theta) = A_{\mathcal{I}_0}^{-1}(b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta) \tag{2.4}$$

   and the optimal dual solution function $y^*(\theta)$ for $\theta \in \Theta_{\mathcal{I}_0}$ is given by

$$y^*(\theta) = y^*(\theta_0). \tag{2.5}$$

3. *The optimal objective function $z^*(\theta)$ or $\theta \in \Theta_{\mathcal{I}_0}$ is given by*

$$z^*(\theta) = c^\intercal A_{\mathcal{I}_0}^{-1}(b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta). \tag{2.6}$$

Proof. See Appendix A.1.1. □

Theorem 1 summarizes the local property of the optimal solutions in a critical region. In particular, each critical region is an open polyhedron in the form of 2.3. The optimal primal solution and the objective function appear to be an affine function of the parameter and the dual solution remains constant within each critical region.

The global property of MPLP over the entire feasible space Θ^* is summarized in the following theorem.

Theorem 2 ([16], p. 126)**.** *The optimal objective function $z^*(\theta)$ is convex and piecewise affine over Θ (and in particular affine in each critical region Θ_i).*

If the optimizer $x^*(\theta)$ is unique for all $\theta \in \Theta$, then the optimizer function $x^*(\cdot)$ is continuous and piecewise affine. Otherwise it is always possible to define a continuous and piecewise affine optimizer function $x^*(\theta) \in \mathcal{X}^*$ for all $\theta \in \Theta$.

Proof. See Appendix A.1.2. □

2.3 Multiparametric Quadratic Programming

Consider the special case of the multiparametric program (2.1) where the objective is quadratic and strictly convex

$$\min_x \tfrac{1}{2} x^\top H x \text{ subject to } Ax \leq b + E\theta. \quad (y) \tag{2.7}$$

where H is assumed to be positive definite[2]

Theorem 3 ([16], p. 132)**.** *Given a parameter θ_0, let $(\mathcal{I}_0, \mathcal{I}_0^c)$ be the optimal partition of the index set $\mathcal{I}$ of (2.2) associated with θ_0. Let $A_{\mathcal{I}_0}, E_{\mathcal{I}_0}$ and $b_{\mathcal{I}_0}$ be, respectively, the submatrices of A, E and subvector of b corresponding to the index set $\mathcal{I}_0$. Let $A_{\mathcal{I}_0^c}, E_{\mathcal{I}_0^c}$ and $b_{\mathcal{I}_0^c}$ be similarly defined for the index set $\mathcal{I}_0^c$. Assume that MPQP (2.7) is neither primal nor dual degenerate for all $\theta \in \Theta$. Denote the critical region that contains θ_0 by $\Theta_{\mathcal{I}_0}$.*

1. *The critical region $\Theta_{\mathcal{I}_0}$ is given by*

$$\Theta_{\mathcal{I}_0} = \{\theta | \theta \in \mathcal{P}_p \bigcap \mathcal{P}_d\} \tag{2.8}$$

where $\mathcal{P}_p$ and $\mathcal{P}_d$ are polyhedra defined by

$$\mathcal{P}_p = \left\{ \theta \;\middle|\; \begin{array}{c} A_{\mathcal{I}_0^c} H^{-1} A_{\mathcal{I}_0}^\intercal (A_{\mathcal{I}_0} H^{-1} A_{\mathcal{I}_0}^\intercal)^{-1} (b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta) \\ < b_{\mathcal{I}_0^c} + E_{\mathcal{I}_0^c}\theta \end{array} \right\}$$

$$\mathcal{P}_d = \{\theta | (A_{\mathcal{I}_0} H^{-1} A_{\mathcal{I}_0}^\intercal)^{-1} (b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta) \leq \mathbf{0}\}.$$

2. *The optimal primal solution function $x^*(\theta)$ for $\theta \in \Theta_{\mathcal{I}_0}$ is given by*

$$x^*(\theta) = H^{-1} A_{\mathcal{I}_0}^\intercal (A_{\mathcal{I}_0} H^{-1} A_{\mathcal{I}_0}^\intercal)^{-1} (b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta) \tag{2.9}$$

and the optimal dual solution function $y^*(\theta)$ for $\theta \in \Theta_{\mathcal{I}_0}$ is given by

$$y_{\mathcal{I}_0}^*(\theta) = -(A_{\mathcal{I}_0} H^{-1} A_{\mathcal{I}_0}^\intercal)^{-1} (b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta), y_{\mathcal{I}_0^c}^*(\theta) = \mathbf{0}. \tag{2.10}$$

3. *The optimal objective function $z^*(\theta)$ for $\theta \in \Theta_{\mathcal{I}_0}$ is given by*

$$z^*(\theta) = \frac{1}{2}(x^*(\theta))^\intercal H x^*(\theta) \tag{2.11}$$

where $x^*(\theta)$ is given by (2.9).

[2]Note that a more general problem $z(x) = \frac{1}{2}x^\intercal H x + h^\intercal x$ can always be transformed into an MPQP of form (2.7) by using the variable substitution $x' = x + H^{-1}h$.

Theorem 3 summarizes the local property of the optimal solutions to the MPQP (2.7) in a critical region. In particular, each critical region is an open polyhedron in the form of 2.8. The optimal primal and dual solutions are affine functions of parameters. The objective function is a quadratic function of parameters.

The global property of MPQP over the entire feasible space Θ^* is summarized in the following theorem.

Theorem 4 ([16], p. 136)**.** *The optimizer function $x^*(\theta)$ is ccontinuous and piecewise affine over Θ (and in particular affine in each critical region Θ_i). The optimal objective function $z^*(\theta)$ is continous, convex and piecewise quadratic over Θ.*

If the MPQP (2.7) is not degenerate, the optimal dual solution $y^*(\theta)$ is continuous and piecewise affine over Θ (and in particular affine in each critical region Θ_i). The optimal objective function $z^*(\theta)$ is continuously differentiable.

The key idea of the proposed multiparametric framework is to have the result of the optimization pre-computed and stored for each parameter in the form as an algebraic function which can be easily evaluated. In other words, the optimal solution of the parameter is explicitly determined and not just implicitly as the result of an optimization problem. To this end, the optimal mappings of the solutions to parameters summarized in Theorem 1-4 constitute the basic results for the applications to the real-time operations under uncertainty.

2.4 Literature Review

In this section, we provide a brief literature review on approaches to multiparametric programs in the form of (2.2) or (2.7). For a more general reference, see [16, 24, 25, 43, 13].

The first method for solving parametric linear programs was proposed by Gass and Saaty [26], and since then extensive research has been conducted to sensitivity and (multi)-parametric linear analysis. One of the first methods for solving MPLPs was formulated by Gal and Nedoma [23]. In [23], a critical region is defined as a subset of the parameter space on which a certain basis of the linear program is optimal. The method proposed in [23] constructs the non-overlapping critical regions iteratively, by visiting the graph of bases associated with the linear program tableau of the original problem.

The geometric approach presented in this thesis was first proposed in [8] for linear programs with a single parameter and generalized to the case with multiple parameters in [15]. In this approach, the definition of critical regions is not associated with the bases but with the set of active constraints. In [14], a simple method for solving MPQPs is presented. The method constructs a critical region in a neighborhood of a given parameter, by using the KarushKuhnTucker conditions for optimality, and then recursively explores the parameter space outside such a region.

CHAPTER 3

PROBABILISTIC FORECASTING OF REAL-TIME LMP AND NETWORK CONGESTION

3.1 Introduction

As more renewable resources are integrated into the power system, and the transmission system operates closer to its capacity, congestion conditions become less predictable and locational marginal prices (LMPs) more volatile. The increased congestion and LMP uncertainties pose significant challenges to the operator and market participants, which motivates us to consider the problem of short-term forecasting real-time LMP/congestion in the presence of generation, demand, and operation uncertainties.

The benefit of accurate LMP and congestion forecasts is twofold. For market participants, accurate forecast of real-time prices is valuable in risk management, bidding strategy development, and demand side participation. The forecasted prices allow market participants to make adjustments in advance to ensure competitive transactions. For system operators, on the other hand, forecast of transmission congestion is important for congestion management and system planning. European transmission system operators, for instance, use Intraday Congestion Forecast (IDCF) to improve real-time security assessment[11]. Intuitively, an LMP forecast should elicit generation participation at times of potential shortage thus alleviating future congestions. Similarly, an LMP forecast can be used for demand response that results in shifting part of the load from peak

to valley.

Currently, some system operators are providing real-time price forecasts. The Electric Reliability Council of Texas (ERCOT) [22] offers a 1-hour ahead real-time LMP forecast, updated every 5 minutes. The Alberta Electric System Operator (AESO) [9] provides two short-term price forecasts with prediction horizons of 2 hours and 6 hours, respectively.

Most LMP forecasting schemes provide only *point forecast*, which gives a single quantity as the prediction. For systems with high levels of uncertainty, a point forecast is rarely accurate, and impacts of prediction error are difficult to quantify. A more attractive alternative is a *probabilistic forecast* that provides a full characterization of the LMP distribution.

Significant technical challenges exist for probabilistic forecasting real-time LMP and congestion. First, reasonably accurate models for real-time dispatch and LMP are needed. Second, network operating conditions and uncertainties need to be incorporated in real time. Finally, the forecasting algorithm needs to be simple and scalable for sufficiently large systems.

These challenges are daunting for market participants who do not have access to network operating conditions and confidential information on bids and offers that influence LMPs. On the other hand, if it is the system operator providing the forecast, as in the case of ERCOT or AESO, the barrier to efficient and accurate forecast is lowered.

3.1.1 Summary of Contributions

In this chapter, we consider the real-time LMP and congestion forecasting problem from an operator perspective. We focus on *probabilistic forecasting* that, at time t, provides the conditional probability distribution at time $t+T$ of the LMP vector and associated congestion, given the system state at time t. Here T is referred to as the prediction horizon which is considered in the range of T from 1 to 6 hours for short-term forecasts.

The key idea behind the proposed approach is the use of multiparametric programming that partitions the uncertainty space into critical regions with each region attached to a unique LMP and congestion pattern. Thus, the problem of probabilistic forecasting reduces to computing the probabilities of random parameters falling in the set of critical regions. When loads or generations (treated as negative loads) are random, their forecasts are incorporated to generate probabilistic LMP and congestion forecasts.

The proposed technique also provides several new features not present in existing methods. For example, it can incorporate system contingency models and allow system constraints to vary with time. The latter feature is relevant because network topology and thermal limits may be changed in real time by the system operator depending on operating conditions. In terms of the generation cost, the proposal can be applied to a linear (or piece-wise affine) function and a quadratic function. Computationally, the proposed method shifts a majority of the computation offline, which significantly reduces online computation cost.

An alternative algorithm that dynamically generates critical regions is also proposed. Because load and stochastic generation processes are physical processes, they are bounded and tend to concentrate around their mean trajectories. Their realizations thus only fall in a few critical regions instead of all over the entire space. By generating the critical regions that contain such realizations, the computation cost is reduced by several orders of magnitude comparing with standard Monte Carlo techniques.

3.1.2 Related Work

Much of the existing work deals with point forecasts of LMP by market participants who do not have access to real-time operating conditions and confidential offers and bids. For these techniques, historical data on LMP, load, and congestions drive the forecasting engine. Literature on these techniques abounds. See [46] and references therein. For probabilistic forecasting techniques by market participants, see approaches in Global Energy Forecasting Competition 2014 [27].

There are several prior studies on LMP/congestion forecasting from the system operator perspective. The proposed technique in [40] employs an online Monte Carlo sampling technique that, for each Monte Carlo sample, solves an optimal power flow (OPF) problem, which is computationally expensive. Monte Carlo technique was also used in [33] where a reduction of the random variable dimension is made using a nonhomogeneous Markov chain model based on a partition of the system state space.

A particularly relevant prior work is [50] where the authors consider the problem of LMP/congestion forecasting from the vantage point of an external observer who has access to publicly available historical data only. Our work, in contrast, considers the forecasting problem from the vantage point of a system operator who has access to the system operating condition at the time of forecasting. In terms of forecasting methodology, the main difference between our approach and that in [50] lies in the different uses of conditioning in evaluating the conditional probability distribution of LMP/congestion.

The authors of [50] introduce and exploit the decomposition of a multi-dimensional load space into critical regions (called system pattern regions) that are estimated using historical data[1]. The work in [50] aims to address the following issue: Given a possible future point L in a multi-dimensional load space, what is the probability distribution of the estimated critical regions that contain L? Since each critical region corresponds to a specific LMP/congestion, the technique in [50] gives a heuristic estimate of the probability distribution of LMP/congestion by conditioning on load L at some future point in time.

In contrast to [50], our objective is to forecast directly the probability distribution of future LMP/congestion, *conditional on the current system operating point*. Because a system operator has access to all private and public information about system conditions, the critical regions are computed exactly via a multiparametric program. This allows us to incorporate load and generation forecasts and obtain the (conditional) probability distribution of future LMP/congestion directly.

[1]The estimated critical regions are therefore random quantities.

3.2 Real-Time Operation Models and Locational Marginal Price

Most wholesale electricity markets [3, 4, 5] consist of day-ahead and real-time markets. The day-ahead market enables market participants commit to buy or sell wholesale electricity one day before operation, and the real-time market balances the difference between day-ahead commitment and the actual real-time demand and production. In this chapter, we focus on real-time operation models. In particular, we consider two real-time markets: one is the energy-only market; the other is the co-optimized energy-reserve market.

3.2.1 Energy Only Market

In the energy-only market, the operator sets generation adjustments by solving a DCOPF problem in which the one-step ahead real-time net load is balanced subject to system constraints [42]. By "net load" we mean the total electrical load plus interchange minus the renewable generation. For simplicity, we assume that each bus has a generator and a load. The DCOPF problem for the operation at time t is defined by the following optimization:

$$\begin{aligned} \min_{g} \quad & c^\top g \\ \text{subject to} \quad & \\ & \mathbf{1}^\top (g - d_t) = 0 && (\lambda_t) \\ & -F^+ \le S(g - d_t) \le F^+ && (\mu_t^+, \mu_t^-) \\ & G^- \le g \le G^+ \end{aligned} \tag{3.1}$$

where

c	real-time generation cost function;
d_t	vector of one-step net load forecast at time t;
g	vector of ex-ante dispatch at time t;
F^+/F^-	vector of max/min transmission capacities;
G^+/G^-	vector of max/min generator capacities;
S	shift factor matrix;
λ_t	shadow price for the energy balance constraint at time t;
μ_t^+/μ_t^-	shadow prices for max/min transmission constraints at time t.

The stochastic generation referenced above can be of any form of renewable integration including renewable energy generation, distributed generation, and demand response. Here we assume that such stochastic generation is non-dispatchable with possible curtailment.

The real-time LMP π_t at time t is calculated from the (dual) solutions of (3.1) as the sum of the energy and congestion prices

$$\pi_t = \lambda_t \mathbf{1} - S^\intercal \mu_t^+ + S^\intercal \mu_t^-. \tag{3.2}$$

From (3.2), we note that the LMP is determined by the marginal generator through λ and the congestion pattern through μ^+ and μ^-. By a congestion pattern we mean a set of congested transmission lines where power flows have reached their limits and a set of uncongested lines. The forecasting of congestion pattern is a sub-problem of the forecasting of LMP.

3.2.2 Joint Energy and Reserve Market

In the joint energy and reserve market, dispatch and reserve are jointly determined via a linear program that minimizes the overall cost subject to operating constraints. In the co-optimized energy and reserve market, system-wide and locational reserve constraints are enforced by the market operator to procure enough reserves to cover the first and the second contingency events. We adopt the co-optimization model in [49] as follows:

$$
\begin{aligned}
\min_{g,r,s} \quad & \textstyle\sum_i \left(c_i^g g_i + \sum_j c_{i,j}^r r_{i,j}\right) + \sum_u c_u^p s_u^l + \sum_v c_v^p s_v^s \\
\text{subject to} \quad & \\
& \mathbf{1}^\intercal (g - d) = 0 && (\lambda) \\
& -F^+ \le S(g - d) \le F^+ && (\mu^+, \mu^-) \\
& \textstyle\sum_i \sum_j \delta_{i,j}^u r_{i,j} + (I_u^+ - I_u) + s_u^l \ge Q_u^l, \forall u && (\alpha_u) \\
& I_u = \textstyle\sum_i \sum_{k \in I_u} S_{ik}(g_i - d_i) && \\
& \textstyle\sum_i \sum_j \delta_{i,j}^v r_{i,j} + s_v^s \ge Q_v^s, \forall v && (\beta_v) \\
& G_i^- \le g_i \le G_i^+ - \textstyle\sum_j r_{i,j}, \forall i && \\
& \hat{g}_{t-1} - \Delta^- \le g \le \hat{g}_{t-1} + \Delta^+ && \\
& 0 \le r \le R^+ && \\
& s_u^l, s_v^s \ge 0, \forall u, v &&
\end{aligned}
\tag{3.3}
$$

where

c_i^g cost for generation at bus i;

$c_{i,j}^r$ cost for reserve type j at bus i;

$c_{u/v}^p$ penalty for reserve deficit of local constraint u or system constraint v;

d vector of net load;

g vector of generation;

$\hat{g}_{t-1}$ vector of generation estimate at time $t-1$;

$r_{i,j}$ generation reserve of type j at bus i;

s^l/s^s vector of local/system reserve deficit;

I_u interface flow for local reserve constraint u;

I_u^+ interface flow limit for local reserve constraint u;

Q^l/Q^s vector of local/system reserve requirement;

R^+ vector of ramp capacities;

Δ^+/Δ^- vector of upward/downward ramp limits;

α_u shadow price for local reserve requirement u;

β_v shadow price for system reserve requirement v;

λ shadow price for the energy balance constraint;

μ^+/μ^- shadow prices for max/min transmission constraints;

$\delta_{i,j}^x$ binary value that is 1 when reserve j at bus i belongs to constraint x.

Similar to the definition of LMP in energy only market, the real-time LMP in the joint energy and reserve market is defined as

$$\pi^{\text{energy}} = \lambda_t \mathbf{1} - S^\intercal \mu_t^+ + S^\intercal \mu_t^- + \sum_u \sum_{k \in I_u} S_{ik} \alpha_u, \tag{3.4}$$

and the reserve clearing price of each reserve product j at bus i is defined as

$$\pi_{i,j}^{\text{reserve}} = \sum_u \delta_{i,j}^u \alpha_u + \sum_v \delta_{i,j}^v \beta_v. \tag{3.5}$$

3.2.3 Parametric Formulation of Economic Dispatch

Our presentation of real-time operation models highlights a *parametric* formulation that treats random elements in the system such as renewable generation, demands, etc., as parameters that vary from time to time and realization to realization.

In the energy only market, the optimization (3.1) can be viewed as a *parametric DCOPF* with parameter $\theta = d_t$. In the joint energy and reserve market, the co-optimization (3.3) can also be viewed as a *parametric DCOPF* with parameter $\theta = (d_t, \hat{g}_{t-1})$ that is realized prior to the co-optimization. This viewpoint plays a critical role in our approach, because each critical region of θ is associated with a unique LMP vector and a particular congestion pattern according to the multiparametric linear programming. When we solve the parametric DCOPF offline, the distribution of real-time LMP and congestion pattern can be directly obtained from the distribution of net load.

In the following, we will use the energy only market model to illustrate the probabilistic forecasting. The energy price and reserve clearing price derived from the co-optimization (3.3) can also be predicted using the same approach.

3.3 Probabilistic Forecasting Algorithm

Probabilistic forecasting, in contrast to point forecasting, aims to provide the probability distribution of a future LMP. In particular, given the estimated system operating point at time t and load and generation forecasts, the probabilistic forecast at time t of the LMP at time $t + T$ is given by the conditional probability distribution $f_{t+T|t}$ of the LMP vector. Entries of the LMP vector are LMPs at individual buses in the system. Since each congestion pattern can be mapped from an LMP vector, we only discuss the probabilistic forecasting of LMP here; the probability distribution of congestion pattern can be obtained from that of LMP.

The key to probabilistic LMP forecasting is to capture spatial and temporal dependencies. Spatial correlations among LMPs arise naturally from the optimization that governs the real-time dispatch. Temporal correlations, on the other hand, are the results of time dependencies in load/generation forecasts. The system randomness may also include occurrences of random contingencies. In this section, we first give a baseline forecasting algorithm. Details on addressing these dependencies are then discussed.

3.3.1 Baseline

The basic idea of the proposed probabilistic forecasting technique is using multiparametric programming analysis to characterize the variation of the real-

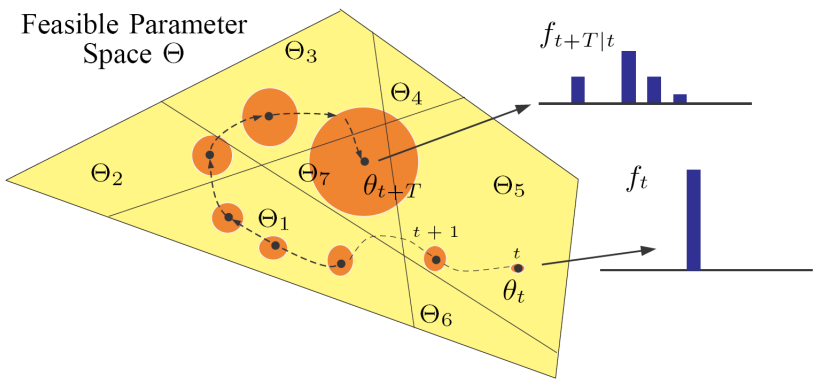


Figure 3.1: Geometric intuition of the proposed algorithm.

time LMP with respect to the random load and generation. By formulating the DCOPF problem (3.1) as a MLP in the form of (2.2), the real-time LMP can be explicitly written as an algebraic function of the random load and generation. The distribution of the LMP vector at a future time can therefore be obtained from the probabilistic forecasts of the stochastic load and generation.

As illustrated in Figure 3.1, load and stochastic generation forecasts in (3.1) are treated as parameters, denoted by $\theta = d_t$. The feasible parameter space Θ is partitioned into critical regions $\{\Theta_1, \cdots, \Theta_7\}$. Within each region Θ_i, the optimal dispatch is associated with the same Lagrange multipliers, and hence a unique LMP vector π_i for all $\theta \in \Theta_i$. Given the network parameters, the MLP solver computes the partition $\{\Theta_i\}$. Correspondences $\{\Theta_i, \pi_i\}$ are then obtained by the Lagrange multipliers and the LMP model (3.2). Note that this computation does not depend on the actual realization of load and generation. Therefore, the computation of the partition and the correspondences can be obtained offline.

Consider now the trajectory of a realization of the random load and generation process θ_t as illustrated in Figure 3.1. Given the realization θ_t and system measurements at time t, we are interested in the conditional probability distri-

bution[2]

$$f_{t+T|t}(i) = \mathbb{P}[\theta_{t+T} \in \Theta_i | \theta_t]. \tag{3.6}$$

As depicted by the shaded circles in Figure 3.1, uncertainties associated with load and generation forecasts increase with time. At time t, the realization of parameter θ_t is known and thus the distribution f_t is an unit vector. But parameter θ_{t+T} may take values from several critical regions where LMP values and congestion patterns are different. Therefore, the forecast probability mass function $f_{t+T|t}$ of $\theta_{t+T|t}$ may have several non-zero elements.

The proposed probabilistic forecasting algorithm involves two parts explained in the following subsections: the computation of critical regions and the estimation of conditional probability distributions, where the former is computed offline and the latter online.

3.3.2 Forecast with Varying Operational Conditions

We describe here a baseline formulation from which critical regions are obtained. We formulate the DCOPF (3.1) used to compute LMP at time $t + T$ as the following right-hand side MLP with the uncertainty parameter θ consisting

[2]We note that this formulation is defined for linear cost functions. If the cost function in the real-time economic dispatch (3.1) is quadratic, the forecast distribution $f_{t+T|t}$ of LMP π_{t+T} can be obtained from the (continuous) distribution of θ_{t+T}, and the uniquely defined affine function of the Lagrangian multiplier vector within each critical region.

of only stochastic load and generation.

$$\begin{aligned} &\min_{g} \quad c^{\intercal} g \\ &\text{subject to} \\ &\begin{bmatrix} \mathbf{1}^{\intercal} \\ -\mathbf{1}^{\intercal} \\ S_{t+T-1} \\ -S_{t+T-1} \\ I_n \\ -I_n \end{bmatrix} g \leq \begin{bmatrix} \mathbf{1}^{\intercal} & \mathbf{1}^{\intercal} \\ -\mathbf{1}^{\intercal} & -\mathbf{1}^{\intercal} \\ S_{t+T-1} & S_{t+T-1} \\ -S_{t+T-1} & -S_{t+T-1} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \theta + \begin{bmatrix} 0 \\ 0 \\ L^{+}_{t+T-1} \\ -L^{-}_{t+T-1} \\ G^{+}_{t+T-1} \\ -G^{-}_{t+T-1} \end{bmatrix} \end{aligned} \tag{3.7}$$

where S_{t+T-1}, $L^{+}_{t+T-1}/L^{-}_{t+T-1}$, and $G^{+}_{t+T-1}/G^{-}_{t+T-1}$ are shift factor matrix, vectors of max/min transmission limits, and vectors of max/min generation capacities at time $t + T - 1$, respectively.

Here we allow *time varying but known system parameters* such as shift factors, flow limits on transmission lines, and limits on generations, restricting uncertainties only to load and generation. The idea is to include deterministically scheduled events in the forecasting problem. Examples include the scheduled changes in network topology [29], and generation capacity and transmission limit.

3.3.3 Forecast in the Presence of Probabilistic Contingencies

The baseline MPLP formulation described in Section 3.3.2 can be extended to include the presence of probabilistic contingencies for unexpected events. For example, a transmission line may be tripped in a storm or generation capacity

reduced due to faults. Such uncertainties in the system parameters need to be handled differently from those associated with the stochastic load and generation.

Because unexpected changes of system configurations are typically small probability events, we assume that there are a total of K possible system configurations at time $t+T$. From historical data, we assume that system configuration k happens with an estimated probability $\hat{p}_k$.

We solve the baseline MPLP (3.7) for each system configuration and obtain critical regions for system configuration k denoted by $\{\Theta_i^{(k)}\}$. By the total probability theorem, the probabilistic forecast of LMP at time $t+T$ is therefore given by

$$f_{t+T|t} = \sum_{k=1}^{K} \hat{p}_k f_{t+T|t}^{(k)}, \tag{3.8}$$

where $f_{t+T|t}^{(k)}$ is the forecast distribution under system configuration k using critical regions $\{\Theta_i^{(k)}\}$ for $k = 1, 2, \cdots, K$.

We illustrate the above idea with an example. Consider the case when at most one of K contingencies can occur between time t and $t+T$. We model the probabilistic contingency as independent tosses of a $K+1$ faced dice where contingency k occurs with probability p_k and no contingency with $p_0 = 1 - \sum_{k=1}^{K} p_k$. We further assume that, once a particular contingency occurs, it remains until time $t+T$. We then solve each MPLP problem (3.7) for all $K+1$ possible system configurations, including the normal condition. The probabilistic forecast

of LMP at time $t + T$ is given by

$$f_{t+T|t} = \begin{cases} f_{t+T|t}^{(k)} & \text{if contingency } k \text{ occurs} \\ \sum_{k=0}^{K} p_k f_{t+T|t}^{(k)} & \text{otherwise} \end{cases} \tag{3.9}$$

where $f_{t+T|t}^{(k)}$ is the forecast distribution under system configuration k with the feasible space $\Theta_{t+T|t}^{(k)}$ for $k = 0, 1, \cdots, K$. Note that system configuration 0 denotes the normal condition.

Having obtained the critical regions, we now consider the problem of computing the conditional distribution of LMP at time $t + T$, given the load and generation forecast at time t.

3.3.4 Probability Distribution Estimation

The estimation of conditional probabilities in (3.6) depends on statistical models of load and generation. Such models can be obtained from either models of the load and stochastic generation process or specific prediction methods used to generate load and stochastic generation forecasts.

As illustrations, we present a directional Gaussian random walk model and an autoregressive (AR) noise model for the random load and generation processes here. It should be noted that any statistical model or prediction method can be applied. The purpose of using these models is to gain insights into the behavior of forecasting performance by taking advantage of some of analytically tractable properties.

A Directional Random Walk Model

We first consider a random walk model of the load/stochastic generation based on a given mean trajectory. Such a model represents a case of minimally informative forecast. Note that the mean trajectory can be any available forecast. For example, a reasonable mean trajectory is the day-ahead load forecast.

Assume that load/stochastic generation θ_t follows a random walk process with a (known) mean trajectory $\bar{\theta}_t$:

$$\theta_t = \theta_{t-1} + \bar{\theta}_t - \bar{\theta}_{t-1} + \epsilon_t, \tag{3.10}$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Given the realization θ_t at time t, the actual load/generation at time $t + T$ is given by

$$\theta_{t+T} = \theta_t + \bar{\theta}_{t+T} - \bar{\theta}_t + \sum_{i=t+1}^{t+T} \epsilon_i. \tag{3.11}$$

Therefore, the distribution of θ_{t+T} conditioning on θ_t is:

$$\theta_{t+T} \sim \mathcal{N}(\bar{\theta}_{t+T|t}, \Sigma_T) \tag{3.12}$$

where $\bar{\theta}_{t+T|t} = \theta_t + \bar{\theta}_{t+T} - \bar{\theta}_t$ is the conditional mean of θ_{t+T} and $\Sigma_T = T\Sigma$ is the cumulative variance within prediction horizon T.

An AR Noise Model

The second model we consider is an AR(1) noise model where we assume the deviation of the load or generation from the expected value is an AR(1) process.

This is a case when the load or stochastic generation is highly structured. In particular,

$$\theta_t = \bar{\theta}_t + a_t, a_t = \phi a_{t-1} + \epsilon_t, \tag{3.13}$$

where $\bar{\theta}_t$ is the (known) mean trajectory, ϕ the parameter of the AR process, and $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Given the realization $\theta_t = \bar{\theta}_t + a_t$ at time t, the noise at time $t + T$ is given by

$$a_{t+T} = \phi^T(\theta_t - \bar{\theta}_t) + \sum_{i=0}^{T-1} \phi^i \epsilon_{t+T-i}, \tag{3.14}$$

and the actual load/generation at time $t + T$ is given by

$$\theta_{t+T} = \bar{\theta}_{t+T} + \phi^T(\theta_t - \bar{\theta}_t) + \sum_{i=0}^{T-1} \phi^i \epsilon_{t+T-i}. \tag{3.15}$$

Therefore, the distribution of θ_{t+T} conditioning on θ_t is:

$$\theta_{t+T} \sim \mathcal{N}(\bar{\theta}_{t+T|t}, \Sigma_T) \tag{3.16}$$

where $\bar{\theta}_{t+T|t} = \bar{\theta}_{t+T} + \phi^T(\theta_t - \bar{\theta}_t)$ is the conditional mean of θ_{t+T}, and $\Sigma_T = \sum_{i=0}^{T-1} \phi^{2i}\Sigma$ the cumulative variance within prediction horizon T.

To sum up, for both models, the conditional probability of θ_{t+T} falling in critical region Θ_i given θ_t is:

$$f_{t+T|t}(i) = \int_{\Theta_i} \frac{\exp\{-\frac{1}{2}(x - \bar{\theta}_{t+T|t})^\intercal \Sigma_T^{-1}(x - \bar{\theta}_{t+T|t})\}}{\sqrt{(2\pi)^n |\Sigma_T|}} dx, \tag{3.17}$$

where $\bar{\theta}_{t+T|t}$ and Σ_T are model associated statistics given in (3.12) and (3.16).

In general, Monte Carlo techniques are necessary to estimate the conditional probability (3.17). To accelerate the sampling process, importance sampling technique is used. In particular, for each critical region Θ_i, instead of drawing values from distribution $\mathcal{N}(\bar{\theta}_{t+T|t}, \Sigma_T)$, we use $\mathcal{N}(\bar{v}(\Theta_i), \Sigma_T)$ where $\bar{v}(\Theta_i)$ is the

mean of all vertices of critical region Θ_i. The estimate of $f_{t+T|t}(i)$ is then given by

$$\hat{f}_{t+T|t}(i) = \frac{1}{N}\sum_{j=1}^{N}\frac{\mathbb{I}_{\Theta_i}(s_j)g(s_j)}{h(s_j)}, \tag{3.18}$$

where samples $\{s_1, \cdots, s_N\}$ are drawn from $\mathcal{N}(\bar{v}(\Theta_i), \Sigma_T)$, and $g(\cdot)$ and $h(\cdot)$ are probability density functions (PDFs) of distribution $\mathcal{N}(\bar{\theta}_{t+T|t}, \Sigma_T)$ and $\mathcal{N}(\bar{v}(\Theta_i), \Sigma_T)$, respectively. Note that the importance distribution $\mathcal{N}(\bar{v}(\Theta_i), \Sigma_T)$ only shifts the mean of the nominal distribution $\mathcal{N}(\bar{\theta}_{t+T|t}, \Sigma_T)$ but keeps the variance the same.

A Quadratic Cost Case

If the cost function in the DCOPF (3.1) is quadratic, the distribution of the LMP π_{t+T} cannot be estimated by the conditional probabilities of θ_{t+T} falling in each critical region. Because the LMP in this case is an affine function of the parameter vector according to Theorem 4.

Here we derive the conditional distribution $f_{t+T|t}$ at time t of the future LMP π_{t+T} at time $t+T$ given the conditional distribution of θ_{t+T}.

By Theorem 4 and the LMP formulation (3.2), for each critical region Θ_i, there exists an affine function of parameters for LMP vectors. Formally, given the distribution of θ, the PDF of the LMP π is summarized in the following theorem.

Theorem 5. *Suppose (3.1) is neither primal nor dual degenerate for all $\theta \in \Theta$ and the distribution of θ is $\mathcal{N}(\mu, \Sigma)$. Denote the critical region partition of Θ by $\{\Theta_1, \cdots, \Theta_N\}$.*

For each critical region Θ_i, there exists U_i and v_i such that

$$\pi_i(\theta) = U_i\theta + v_i, \text{ for all } \theta \in \Theta_i.$$

Let $\bar{S}_i$ be the submatrix of S corresponding to the active transmission constraints in the optimal partition associated with Θ_i. If $\begin{bmatrix} \mathbf{1} & \bar{S}_i^\intercal \end{bmatrix}$ has full column rank for all i, the distribution of LMP π is given by

$$f(\pi) = \sum_{i=1}^{N} \mathbb{I}_{\Pi_i}(\pi) f_i(\pi), \tag{3.19}$$

where Π_i is the image of $\pi_i(\cdot)$ and $f_i(\pi)$ the PDF of $\mathcal{N}(U_i\mu + v_i, U_i\Sigma U_i^\intercal)$.

The closed form expression of $f_{t+T|t}(\pi)$ follows Theorem 5 immediately. Given the conditional distribution $\mathcal{N}(\bar{\theta}_{t+T|t}, \Sigma_T)$ of θ_{t+T}, the conditional distribution $f_{t+T|t}$ at time t of the future LMP π_{t+T} at time $t+T$ is given by

$$f_{t+T|t}(\pi) = \sum_{i=1}^{N} \mathbb{I}_{\Pi_i}(\pi) f_{t+T|t,i}(\pi), \tag{3.20}$$

where $f_{t+T|t,i}(\pi)$ is the PDF of $\mathcal{N}(U_i\bar{\theta}_{t+T|t} + v_i, U_i\Sigma_T U_i^\intercal)$.

In summary, the proposed algorithm treats load/generation as parameters, formulates the DCOPF (3.1) as a MPP (2.2), determines the critical regions, and computes the conditional distribution of LMP and congestion using load/generation forecasts and the real-time operation conditions. These system conditions, such as transmission rate, generation capacity, and network topology, are allowed to vary with time but known to the operator at the time of forecasting. Contingency models are also incorporated in the proposed technique.

3.4 Online Forecasting via Dynamic Critical Region Generation

A limiting factor in the proposed technique above is the computational cost associated with the multiparametric programming. Since the solution structure is characterized by critical regions, all critical regions that partition the parameter space have to be calculated. Although such computation can be made offline, it may not be computationally tractable for large systems because the number of critical regions may grow exponentially with the number of constraints.

In this section, we propose an online Monte Carlo technique, referred to as dynamic critical region generation (DCRG), that significantly reduces the computational cost. The idea is to take advantage of the fact that, in practice, random load and generation processes are bounded and tend to concentrate around their mean trajectories. As a result, a small fraction of critical regions represents the overwhelming majority of observed critical regions. When a parameter falls in a critical region that was visited before, the Lagrange multipliers that are used to compute LMPs can be obtained directly from the affine mappings associated with that critical region, without solving the DCOPF (3.1).

The key idea of DCRG, therefore, is to compute on demand the critical region and the associated coefficients of the affine mapping of parameter to the LMP. This computation, fortunately, is no more than elementary matrix inversions and multiplications. The computation procedure is given by the proof of Theorem 1 for the linear cost and the proof of Theorem 3 for the quadratic cost.

In applying DCRG to LMP forecasting using online Monte Carlo simulation, we generate samples of random load or generation, either based on load/generation models or from historical data. Instead of directly simulating the real-time market operation as in [40], we check if the generated sample falls into a critical region that has been used before. If it does, the LMP can be generated using the affine mapping of Lagrangian multipliers directly without solving the DCOPF (3.1). If the parameter does not belong to any critical region in the database, a DCOPF is solved and a critical region containing the parameter is computed according to (2.3) for linear cost case or (2.8) for quadratic cost case in Theorem 3.

We present in this section a new methodology of probabilistic forecasting and online simulation of real-time electricity market. In particular, we are interested in obtaining conditional probability distributions of future LMPs, power flows, dispatch levels, and congestion pattern from sample paths of random processes of stochastic parameters such as load and generation processes. These sample paths can be generated via Monte Carlo simulation based on stochastic models or by sampling historical traces.

Our approach is one of online learning that acquires sequentially a set of solutions that most frequently appear in Monte Carlo simulations, which allows us to avoid explicit computations of DCOPF solutions. In particular, we borrow the notion of dictionary learning to explain the ideas behind the proposed online learning approach to forecasting. Widely used in the signal processing community, dictionary learning refers to acquiring a dictionary of signal bases to represent a rich class of signals using words (atoms) in the dictionary [36, 44].

There are two components of the online learning approach. One is the learning of the underlying stochastic model of the parameter process, and the other is the learning of the collection of critical regions that characterizes the solution structure of parametric DCOPF. Since there is an extensive literature on the former, we focus here on the problem of learning the structure of parametric DCOPF.

Analogues to dictionary learning in signal processing, the learning process here is also acquiring a dictionary whose words (or atoms) are critical regions. In particular, each word is associated with the affine function that maps the parameter to the solution of MPLP/MPQP. Therefore, if we treat a realization of the parameter process as a sentence, the dictionary allows us to translate a sentence in the language of system parameters to one in the language of LP/QP solutions.

The ODL process therefore includes (i) checking if a new parameter θ has already been learned in the past. If not, (ii) construct a new entry in the dictionary by computing the critical region that contains θ. For (ii), the construction of the dictionary is given by Theorem 1. The detailed algorithm is summarized in Algorithm 1.

3.5 Evaluation

In this section, we present simulation results to compare performances of the proposed probabilistic forecasting algorithm with some benchmark methods.

Algorithm 1 Online Dictionary Learning for Critical Regions

given the mean trajectory $\{\bar{d}_t\}_{t=1}^T$ of the net load and associated (forecast) distributions $\{\mathcal{F}_t\}_{t=1}^T$
initialize the critical region dictionary C_0 using the mean trajectory
for $m = 1, \cdots, M$ **do**
  for $t = 1, \cdots, T$ **do**
    Generate sample d_t^m and let $\theta_t^m \triangleq (d_t^m, g_{t-1}^m)$.
    Search $\mathcal{C}_{t-1}^m$ for critical region $C(\theta_t^m)$.
    if $C(\theta_t^m) \in \mathcal{C}_{t-1}^m$ **then**
      Compute g_t^m from the affine mapping $g_{C(\theta_t^m)}^*(\theta_t^m)$.
    else
      Solve g_t^m from DCOPF (3.1) using θ_m^t, compute $C(\theta_t^m)$, and update $\mathcal{C}_t^m = \mathcal{C}_{t-1}^m \cup \{C(\theta_t^m)\}$.
    end if
  end for
end for

We first show results of a 3-bus system with a linear cost function to gain insights into the behavior of the proposed algorithm under various scenarios. Simulations using the IEEE 118-bus system with a quadratic cost function are then presented to demonstrate the scalability of the proposed algorithm and the effectiveness of the online heuristic approach given in Section 4.5.

3.5.1 Benchmarks and Performance Measure

We compared the proposed techniques with some existing benchmarks for forecasting and computation performance. Since, to our best knowledge, there is no probabilistic forecasting techniques for the operator in the literature, we used the direct Monte Carlo simulation method proposed in [40] as the probabilistic forecasting benchmark. It should be noted that the direct Monte Carlo simulation approach generates exactly the same probabilistic forecast as the pro-

posed technique. We also included comparisons of the proposed technique with two well known point forecasting methods to illustrate the performance gain. In particular, the deterministic prediction uses the mean trajectory of load/stochastic generation $\bar{\theta}_{t+T}$ to calculate LMP and congestion pattern at time $t+T$. The certainty equivalence prediction incorporates measurements at time t and uses the conditional mean trajectory $\bar{\theta}_{t+T|t}$.

Before presenting numerical results, we introduce a performance evaluation metric of probabilistic forecasts. The LMP[3]/congestion pattern is a discrete random vector. The probabilistic forecast of such a random quantity belongs to the so-called categorical forecast, and its performance is measured by the consistency as well as the statistical concentration of the forecast. A standard metric for this type of forecast is the Brier Score (BS) [17] that measures the average distance (2-norm) between the forecast distribution $f_{t+T|t}$ and the point mass distribution at the realized random variable π_{t+T}. Specifically,

$$\mathrm{BS}(f_{t+T|t}) = \mathbb{E}\|f_{t+T|t} - \delta(\pi_{t+T})\|^2 \tag{3.21}$$

where the expectation is taken over all randomness between time t and $t+T$. In (3.21), $f_{t+T|t}$ is the conditional probability vector whose ith entry is given by $f_{t+T|t}(i) = \Pr(\pi_{t+T} = \pi_i)$, and $\delta(x)$ is the unit vector that is one at entry x and zero elsewhere. This score ranges from 0 for a perfect forecast to 2 for the worst possible forecast.

Since BS is a succinct formula to measure the overall performance in terms of uncertainty, reliability and resolution, we also provide a more intuitive assessment — reliability diagram. Reliability diagram is a graph of the observed

[3]Note that the discreteness of LMP is only for linear or piece-wise affine cost functions.

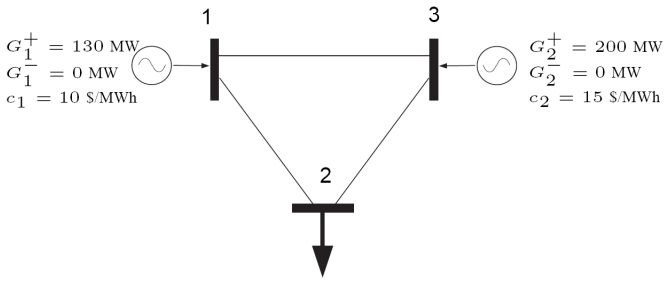


Figure 3.2: A 3-bus system.

frequency of an event plotted against the forecast probability of an event. It measures how closely the forecast probabilities of an event correspond to the actual chance of observing the event.

3.5.2 Example 1: A 3-Bus System

Consider a 3-bus system as depicted in Figure 3.2. Generator incremental costs and capacity limits are presented in the figure. All lines are identical with the maximum capacity of 100 MW. The prediction horizon is one hour (12-step ahead). Since the direct Monte Carlo simulation approach has the same forecasting performance as the proposed algorithm, we only present results of the proposed algorithm in this case study.

Baseline

We first evaluated the baseline algorithm with the two load models. Note that only the load at bus 2 was stochastic, which gave the one dimensional parametric linear program.

Table 3.1: Critical regions, LMPs and congestion patterns for the baseline[4].

	Critical Region	LMP	Congestion
1	(0, 130)	(10, 10, 10)	(0, 0, 0)
2	(130, 170)	(15, 15, 15)	(0, 0, 0)
3	(170, 200)	(10, 20, 15)	(1, 0, 0)

Table 1 shows the critical regions of the parametric linear program, and the associated LMPs and congestion patterns. In this case, the parameter load d at bus 2 was partitioned into three segments $\mathcal{D} = \{(0, 130), (130, 170), (170, 200)\}$.

We used a straight line ranging from 110 MW to 190 MW with 2 MW increments as the mean trajectory of load $\bar{d}_t$. The coefficient ϕ in AR(1) noise model was set at 0.9. The independent noise sequence ϵ_t in both models followed the standard normal distribution, *i.e.*, $\mathcal{N}(0, 1)$. Monte Carlo simulations were used to obtain estimated BSs in Figure 3.3.

From Figure 3.3, we observed that the proposed probabilistic forecasting algorithm "Alg-P" consistently outperforms the deterministic "Alg-D" and certainty equivalence "Alg-C" predictors in both load models. The superior performance of the proposed technique in these two extreme models (a minimally informative model and a highly structured model) shows its capability of incorporating different load forecasting methods and its forecasting power. Two interesting phenomena are worth closer examinations. First, for both load models, peaks occurred at the boundaries of two neighboring critical regions when the mean load $\bar{d}_t$ is 130 at $t = 10$ and 170 at $t = 30$. At the boundary point, the probability of d_{t+T} falling in either the left or the right critical region was the

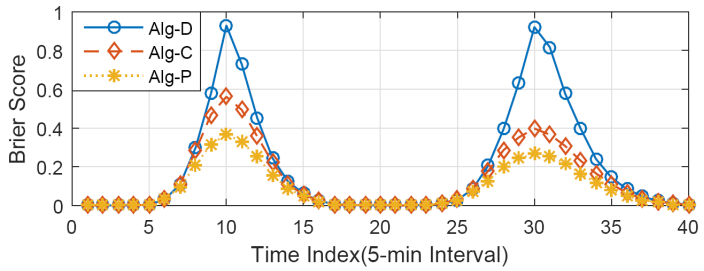
[4]In Table 3.1, the triple for LMP contains price values at bus 1-3. For congestion, the triple summarizes status of line 1-2, line 1-3, and line 2-3, respectively, where "1" indicates positive congestion (the flow reaches the line limit in the positive direction), "−1" negative congestion, and "0" no congestion.

same. Roughly half of the time Alg-D predicted the LMP correctly, the other half was completely wrong. From the definition in (3.21), the BS of Alg-D should be 1.

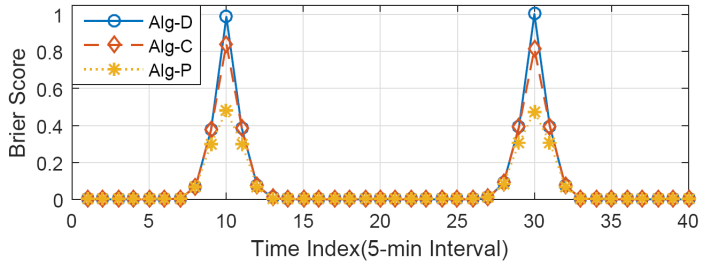
Second, a more subtle point, the scores for the random walk model showed a slight asymmetry with respect to boundaries of neighboring critical regions: the BSs of Alg-P and Alg-C at the second peak were lower than that of the first peak, and the ranges of non-zero score neighborhood of the second peak (from time 24 to 39) of all three algorithms were wider than that of the first peak (from time 5 to 16). This phenomenon arose primarily from the process of generating the sample trajectory of load. Since the entire sample trajectory was generated at once, the deviation $|d_t - \bar{d}_t|$ from the mean grows over time which leads to a bigger variance of the time crossing the second boundary 170 than that of the time crossing the first boundary 130. In other words, comparing to the probability of crossing the first boundary 130 at $t = 10$, the probability of crossing the second boundary 170 at $t = 30$ is lower, but the probability in its neighborhood is higher. Therefore, the scores of Alg-P and Alg-C at the second peak were lower, and the ranges of non-zero score neighborhood bigger.

Forecast with probabilistic generation outage

We then considered the case of a generating unit outage with a partial loss of capacity, assuming that the maximum capacity of the generator at bus 1 can be reduced to 100 MW with probability p. Other settings were the same as those in the first scenario.



(a) Random walk model



(b) AR(1) noise model

Figure 3.3: Impact of load statistical models.

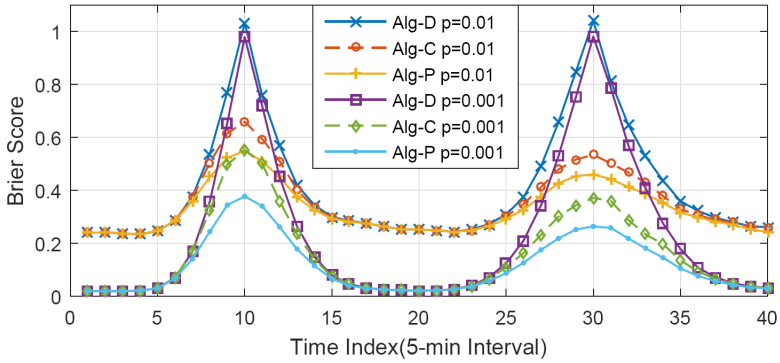


Figure 3.4: Impact of outage frequency p.

The critical regions under this configuration became $\mathcal{D}^{\text{out}} = \{(0, 100), (100, 200)\}$. To predict a future price, we considered all critical regions $\{\mathcal{D}, \mathcal{D}^{\text{out}}\}$ that load d_{t+T} may fall in, where $\mathcal{D}$ refers to the critical regions in Table 3.1 under normal conditions. For the outage frequency p, we chose two levels: $p = 0.01$ and $p = 0.1$. The random walk model was adopted to generate load profiles. Benchmark algorithms also took contingencies into consideration for a fair comparison.

From Figure 3.4, the behaviors of all three algorithms with the same outage frequency p were similar to that in the first scenario. Boundary effects and peak asymmetries were observed with contingencies as well. Compare the performances of each algorithm with different outage frequencies: the smaller the outage rate the better the forecast.

3.5.3 Example 2: IEEE 118-Bus System

The IEEE 118-bus system was used to show the scalability of the proposed technique and the effectiveness of the heuristic algorithm described in Section 4.5. The simulation results were focused on the complexity comparison between the proposed techniques and the probabilistic forecast benchmark.

We introduced 12 wind generators (roughly 10% the number of buses) with $10 \sim 20\%$ renewable penetration[5]. The wind generators were located at bus 25, 26, 90, 91, 100, 103, 104, 105, 107, 110, 111, and 112. The selection of these locations were intended to represent two wind farms, the small one has 2 wind generators, and the large one has 10 wind generators concentrated on a few neighboring buses. All wind generators were assumed to be identical with the maximum capacity of 110 MW. Denote the wind generation space by the hypercube $\mathcal{W} \in \mathbb{R}^{12}$. We imposed the maximum capacity of 100 MW on transmission line 8, 126, and 155. The load profile, generator capacities and cost functions, and line and bus labels were referred to as in "case118" in MATPOWER [51].

[5]By $x\%$ renewable penetration we mean the mean value of total wind generation is $x\%$ of the total electricity load (4242 MW in this system). Note that load was assumed to be deterministic in this case.

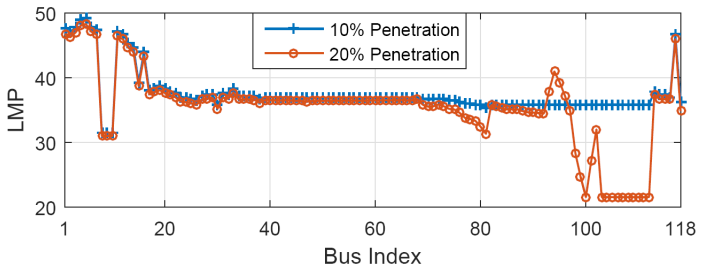


Figure 3.5: LMPs at 10% and 20% renewable penetration levels.

Note that the cost function in this system was quadratic, thus the LMP was an affine function of the wind production within each critical region.

The mean trajectory $\bar{w}(i)$ of each wind generator i was assumed to be linear, for $i = 1, \cdots, 12$. In particular, the trajectory starts from 10% penetration level, *i.e.*, $\bar{w}_0(i) = 35.35$, at time 0, and ends at 20% penetration level, *i.e.*, $\bar{w}_{10}(i) = 70.70$, at time 10. The increment was assumed to be constant, *i.e.*, 3.535 MW. LMP values at 10% and 20% penetration levels are given in Figure 3.5. We observed that higher renewable penetration reduced LMP at most buses, but raised the LMPs at bus 93, 94, 95, and 96. The reason of such nonintuitive increase was the congestion of transmission line 155 caused by the increased wind production from the big wind farm. The random walk model was used for the stochastic generation profile. The distribution of the independent noise process ϵ_t was set to be the standard multivariate Gaussian.

Results for the 10-step ahead prediction at time $t = 0$ are provided in Fig. 3.6. The predicted marginal distribution of LMP at bus 49 exhibits a Gaussian distribution as it is well fitted by the Gaussian approximation with the sample mean and sample variance as distribution parameters. However, such Gaussian characteristics were not observed on the other three buses. According to the distributions of LMP at bus 94 and 100, the extreme values of LMP occasionally

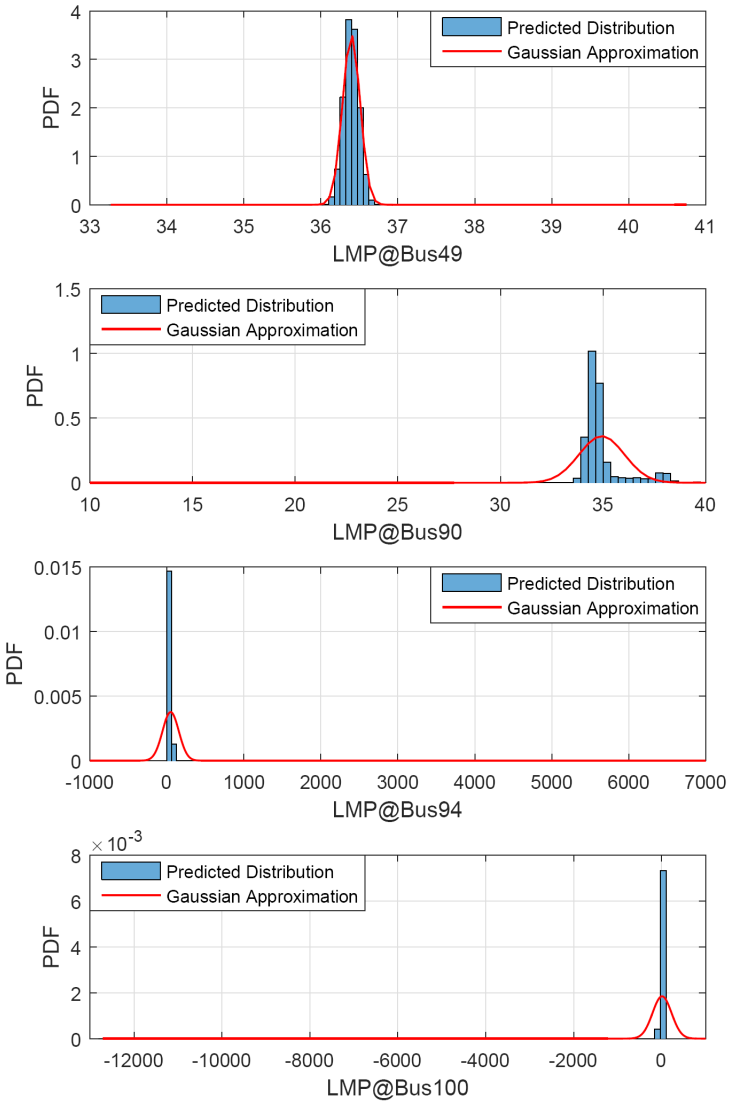


Figure 3.6: Sample predicted LMP distributions and Gaussian approximations.

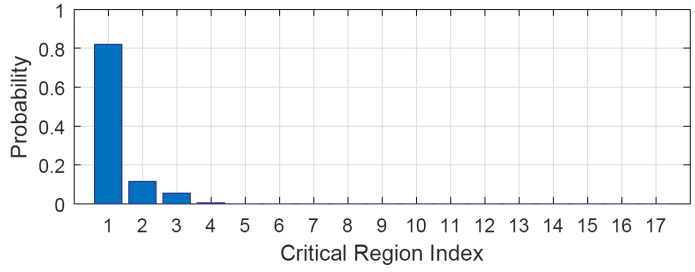


Figure 3.7: Conditional distribution of observed critical regions.

appeared as spikes, which were caused by the network congestion.

In the following, we focus on the complexity performance of the proposed forecasting techniques. Theoretically, there were at most 2^{115} critical regions, because the constraints in the DCOPF (3.1) for this example included 1 energy balance constraint, 6 transmission constraints, and 108 generator capacity constraints. For the given hypercube $\mathcal{W}$ of the parameter space, there were in total 273 critical regions. But for the generated 10,000 samples at time 10, only 17 critical regions were observed and more than 99% samples fell in 3 critical regions, as shown in the distribution over the observed 17 critical regions in Figure 3.7.

Instead of exploring the entire wind production space offline, we implemented the online forecasting algorithm DCRG given in Section 4.5. Figure 3.8 shows the comparison of computational cost between the proposed DCRG algorithm, labeled by "Alg-DCRG", and the direct Monte Carlo simulation, labeled by " Alg-MC". Both algorithms present approximately linear growth in the logarithm scales. But the proposed DCRG algorithm provided more than three orders of magnitude reduction in the number of DCOPF computations required in the simulation.

Finally, we provide the computation time comparison for the three probabilistic forecasting techniques in Table 3.2. All computational times were evaluated by implementing the algorithms in Matlab environment with the default "quadprog" solver and an external MPT3 [30] toolbox on a desktop with an Intel Core i7-3770 CPU at 3.4 GHz and 8 GB memory. No attempts were made to optimize the efficiency of the algorithms and their simulations. From Table 3.2, we

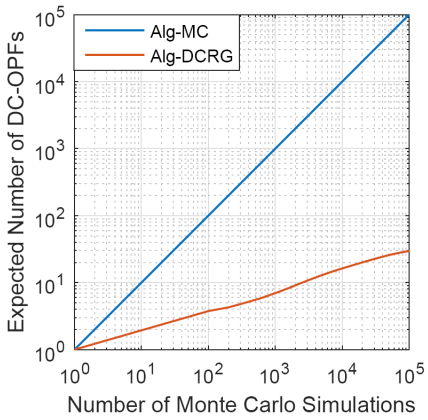


Figure 3.8: The expected number of DCOPF computations versus the total number of Monte Carlo simulations.

Table 3.2: Computation time (in seconds) for 10,000 samples.

	Offline	Online	Total
Alg-P	160.60	23.32	183.92
Alg-DCRG	—	23.59	23.59
Alg-MC	—	52022.57	52022.57

can conclude that the direct Monte Carlo simulation approach Alg-MC failed to meet the time constraint for *real-time* LMP forecasting, since 10,000 samples took more than 14 hours to generate the distribution. The proposed techniques, on the other hand, only took less than half a minute in the online computation, demonstrating the efficiency for the real-time LMP forecasting.

CHAPTER 4

STOCHASTIC MULTI-AREA INTERCHANGE SCHEDULING

4.1 Introduction

Since the restructuring of the electric power industry, ISOs and RTOs have been created to operate power grids defined by certain geographical boundaries. Within each control region, a centralized market is administered by the operator who collects supply offers and demand bids, determines market clearing prices, and settles auctions for producers and consumers. On the boundary, trades between two regions are jointly determined by both operators to ensure their physical deliveries. Such inter-regional trades enable market participants to buy electricity from one region and sell it to the other, and allow low-cost external resources to compete with internal resources to serve consumers.

If there were no trading barrier across neighboring areas, there would have been a seamless market as a whole, approximating a single larger market. In practice, however, the so-called seams between neighboring operating regions do exist, which is manifested by the counter-intuitive flows that export power from high cost to low cost regions. Another symptom of seams is the under-utilization of the interface transfer capacity when more power could have been exported from a low cost region to a high cost one.

Ideally, seams problem can be eliminated completely by jointly optimizing the overall cost of interconnected regions through a DCOPF algorithm. There

is an extensive literature on solving the decentralized multi-area economic dispatch problem as summarized briefly in Section 4.1.1. These approaches, however, have not been adopted in practice for economic, political, and technical reasons.

The current industrial approach is a market-based solution that sets the net interchange[1] based on bids and offers submitted by market participants under a set of complicated rules and procedures. In [47], the analysis of the seams issues between New York ISO and ISO New England shows that the economic loss due to seams for the New York and New England customers is estimated at the level of $784 million from 2006 through 2010.

The authors of [47] point out that the latency between the scheduling of interchange and the actual power delivery is a major cause of inefficiency. Typically, the net interchange is set by a market clearing process in which external transaction offers and bids are submitted far ahead of the time of power transfer. Consequently, the interchange determined in advance may not reflect the actual system conditions. This situation is likely to be exacerbated with the greater integration of renewables.

A second factor that causes inefficiency is the lack of coordination between neighboring ISOs in their respective clearing processes of bids and offers from external market participants. This suggests that considerable gain in performance may be realized by a process that schedules the interchange based on the minimization of overall system cost.

[1] By the net interchange we mean the total power transferred from one region to another over the boundary.

4.1.1 Related Work

There have been extensive studies on the seams issue. In this chapter, we do not consider inefficiencies arise from market designs; we focus instead on optimizing the interchange schedule. Techniques aimed at improving interchange efficiency can be classified into two categories. The first aims to optimize the overall interconnected system in a decentralized fashion. In particular, the optimal interchange schedules can be obtained from the multi-area OPF problem. For broadly related work on multi-area OPF, see [47, 35, 20, 48, 12, 18] and references therein.

Mathematically, the optimal interchange can be obtained from the multi-area OPF problem using various decomposition techniques. A general approach is based on the principle of Lagrangian Relaxation (LR) that decomposes the original problem into smaller subproblems. Some of the earliest approaches include the pioneer work of Kim and Baldick [35] and Conejo and Aguado [20] that predate the broad deregulation of the electricity market in the U.S. In general, decentralized OPF techniques typically require iterations between interconnected regions where one ISO uses intermediate solutions from the other to solve its own dispatch problem. Although the convergence of such techniques is often guaranteed under the DCOPF formulation, the number of iterations can be large and the practical cost of communications and computations substantial. An exception is the recent marginal decomposition technique [48] that is shown to converge in a finite number of iterations.

The growth of renewable integration has brought new attention to uncer-

tainty in seams. Both stochastic optimization and robust optimization approaches have been considered recently. In particular, the authors of [10] formulate a two-stage stochastic market clearing model for the multi-area energy and reserve dispatch problem. The solution to the stochastic optimization is obtained based on scenario enumerations, which requires a prohibitively high computation effort. In [37], the day-ahead tie-flow scheduling is considered in the unit commitment problem under wind generation uncertainty. Specifically, a two-stage adaptive robust optimization is formulated with the goal of minimizing the cost of the worst-case wind production. The solution is given by the column-and-constraint generation algorithm. In [21], an adjustable interval robust scheduling of wind power for day-ahead multi-area energy and reserve market clearing is proposed. The uncertainty of wind farms is represented by predefined intervals and the clearing model is formulated as a mixed integer quadratic programming problem.

The main issue of this category approach is the elimination of arbitrage opportunities for the external market participants. Since operators cannot trade with each other directly, market participants facilitate trades between control areas. The multi-area economic dispatch approach thus cannot be implemented under the current regulation.

The second category includes the current industrial practices based on the so-called proxy bus approximation [19, 47]. The proxy bus is a trading location at which market participants can buy and sell electricity. In [19], a coordinated interchange scheduling scheme is proposed for the co-optimization of energy and ancillary services. In [47], an interchange scheduling technique, referred

to as Tie Optimization (TO), is proposed. An alternative, referred to as Coordinated Transaction Scheduling (CTS), is also proposed. Built upon TO, CTS is an enhancement of the market-based solution with higher scheduling frequency and more tightly coordinated scheduling between two ISOs. As a state-of-the-art scheduling technique, CTS uses economic argument on supply and demand functions exchanged by the neighboring operators. When there is only a single interface in a two-area system, such functions can be succinctly characterized, and the exchange is only made once; the need of iterations among operators is eliminated. Our approach is also based on the same economic argument with the innovation on incorporating system and operation uncertainty and the generalization to multi-area multi-interface systems.

4.1.2 Summary of Contributions

In this chapter, we consider the optimal scheduling of interchange *in the presence of system and operation uncertainty*. To this end, we propose a two-stage stochastic optimization formulation aimed at minimizing the expected overall system cost. The proposed optimization framework takes into account random fluctuations of load and renewable generation in the system. To solve the intractable stochastic optimization, especially considering continuous random variables, we start from the simplistic two-area single-interface system to gain some fundamental insights and then generalize to the general multi-area multi-interface systems.

The main contribution of this chapter is threefold. First, we present an ap-

proach to transfer the stochastic interchange scheduling problem for the two-area single-interface system into an equivalent but deterministic optimization. This transformation allows us to generalize the deterministic TO solution by intersecting the *expected* demand and supply functions of interchange, therefore avoiding repetitive computation and iterative information exchange between operators.

Second, we generalize the single interface scheduling problem to the case involving a network of operating areas, each having multiple interfaces with its neighbors. Specifically, we consider two types of interface scheduling: (i) the synchronous scheduling and (ii) the asynchronous scheduling. The former requires all areas operated under the same scheduling clock whereas the latter allows every pair of operating areas setting their interfaces independently of others. To our best knowledge, there is no existing results on this problem in the open literature.

The challenge for the generalization arises from the fact that the interfaces cannot be succinctly characterized by a pair of expected demand and supply functions — an essential property underlying the approach for the single interface stochastic scheduling. When multiple interfaces are involved, the simple idea of equating expected demand and supply functions is not applicable and there is no simple notion that the intersection of demand and supply curves gives the social welfare optimizing interchange.

Finally, to solve the optimal stochastic interchange scheduling in multi-area multi-interface system, we present a new scheduling technique based on the

classical idea of coordinate descent method. The main idea is to iterate over all interfaces, one at a time, minimizing the overall system cost under uncertainty. In each iteration, a single interface optimization problem is solved by projecting the current solution to a particular coordinate representing a particular interface flow. The optimality and convergence are guaranteed under nominal assumptions for both synchronous and asynchronous algorithms.

4.2 Proxy Bus Representation

In practice, coordination between neighboring control regions and markets is typically carried out through the use of a proxy bus mechanism. According to [28], the proxy bus mechanism is utilized by all existing LMP based markets for representing and valuing interchange power. For interchange scheduling, a proxy bus is the location to which generation in the neighboring area is *assumed* be dispatched up and down in response to the change of interchange schedule.

Depending on the number of proxy buses used by operators to model imports and exports, the proxy bus representation can be categorized into two types: (i) the single proxy bus system and (ii) the multiple proxy bus system.

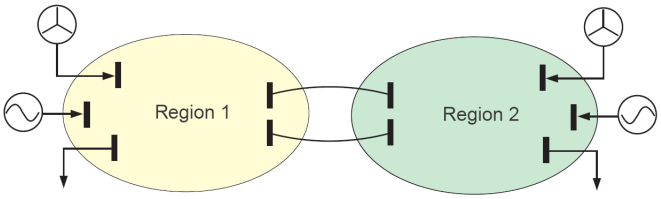


Figure 4.1: A two-area single-interface system.

4.2.1 Single Proxy Bus Representation

In a single proxy bus system, all interchange scheduled with an adjacent dispatch area is assumed to result in changes in net generation at the location of the proxy bus [28]. As an example, Figure 4.1 illustrates a power system consisting of two independently operated subsystems, each having its own internal load, generation and renewable integration. In the *single proxy bus system*, each operator selects *one* proxy bus to represent the location of import or export in the neighboring region. Specifically, as shown in Figure 4.2, ISO 1 assumes a withdrawal q at proxy bus p_1 and ISO 2 an injection with the same quantity q at proxy bus p_2. In this example, the direction of the net interchange q is defined from region 1 to region 2. Note that a proxy bus can be a physical location or a virtual abstraction. For simplicity, we assume all proxy buses are physical locations throughout this thesis.

4.2.2 Multi-Proxy Bus Representation

In a multiple proxy bus system, operators use two or more proxy buses for representing transactions with adjacent control areas or dispatch regions. The multi-proxy bus systems are referred to as situations where multiple proxy

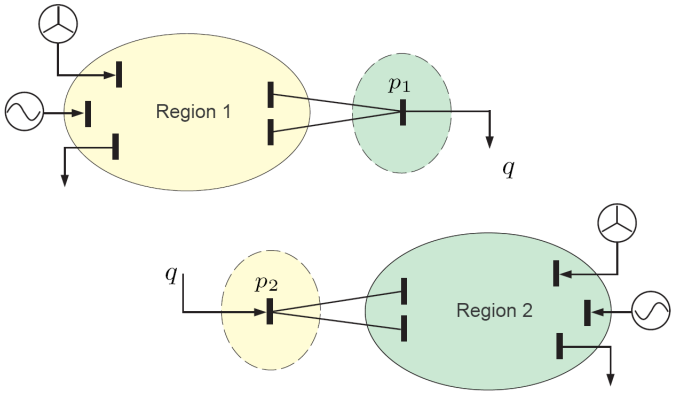


Figure 4.2: The single proxy bus representation.

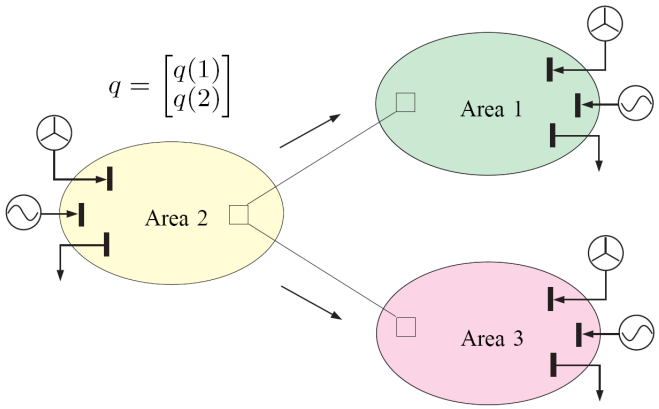


Figure 4.3: A multi-proxy bus representation.

buses are defined within a single control area, and situations there is a single proxy bus in each control area but there are multiple adjacent control areas.

We use an example to illustrate the multi-proxy bus system in Figure 4.3 where three proxy buses, indicated by boxes, are defined in a 3-area 2-interface interconnected system. The interchange vector q consists of two interface flows, $q(1)$ and $q(2)$, with fixed directions indicated by the arrows in Figure 4.3. Area n maintains a regional interchange vector q_n that describes its own interface flows. For convenience, we assume the direction of each interface flow in the regional interchange vector being *outbound*. In this example, the interchange vector for

area 1 is $q_1 = -q(1)$, area 2 $q_2 = (q(1), q(2))$ and area 3 $q_3 = -q(1)$.

The interchange scheduling is to determine the direction and volume of all interface flows that minimize the overall generation cost subject to the power balance constraint and generation and transmission constraints *under the proxy bus model*. This should be distinguished with the problem of multi-area economic dispatch in which the optimal tie line flows and regional generation dispatch are optimized *without network approximation*. Since the proxy bus representation is an approximation, the optimal interchange in the proxy bus system may not be optimal to the original system. In general, the optimal interchange via proxy bus representation is strictly suboptimal when it is compared with multi-area OPF solutions.

4.3 Deterministic Single Interface Scheduling

In this section, we consider the *deterministic* interchange scheduling problem for the *two-area single-interface* system under the single proxy bus mechanism in Figure 4.2.

4.3.1 Problem Formulation

The interchange scheduling problem under the proxy bus model can be formulated as minimizing the generation costs of both regions subject to the power

balance, (internal and interface) transmission and generator constraints. For simplicity, it is assumed in this chapter: (i) the system is lossless, and (ii) the regional cost function of generation $c_i(\cdot), i \in \{1, 2\}$, is quadratic and strictly convex. In the single proxy bus system, the net interchange can be modeled explicitly as an additional scalar variable in the optimization as follows:

$$
\begin{aligned}
\min_{q, g_1, g_2} \quad & C_1(g_1) + C_2(g_2) \\
\text{subject to} \quad & \mathbf{1}^\intercal (d_1 - g_1) + q = 0 && (\lambda_1) \\
& \mathbf{1}^\intercal (d_2 - g_2) - q = 0 && (\lambda_2) \\
& S_1(d_1 - g_1) + s_1 q \leq F_1 && (\mu_1) \\
& S_2(d_2 - g_2) - s_2 q \leq F_2 && (\mu_2) \\
& q \leq Q && (\mu_q) \\
& g_1 \in \mathcal{G}_1 \\
& g_2 \in \mathcal{G}_2
\end{aligned}
\tag{4.1}
$$

where

$\mathbf{1}$	vector of ones with compatible dimensions;
$C_n(\cdot)$	generation cost function for area n, assumed to be in the form $C_n(g_n) = \frac{1}{2} g_n^\intercal H_n g_n + l_n^\intercal g_n$, where H_n is positive definite;
d_n	vector of forecasted net load including demand and renewable generation for area n;
g_n	vector of dispatches for area n;
q	net interchange from area 1 to area 2;
s_n	shift factor vector of proxy bus to internal transmission lines in area n;
F_n	vector of internal transmission limits for area n;
Q	interface capacity;
S_n	shift factor matrix of buses in region n to internal transmission lines in area n;
$\mathcal{G}_n$	generator constraints for area n;
λ_n	shadow price for power balance constraint in area n;
μ_n	shadow prices for transmission constraints in area n;
μ_q	shadow price for the interface capacity constraint.

The problem (4.1) is a centralized formulation for determining the optimal interchange between area 1 and 2. Such an optimization requires a coordinator who have full access to all related information of both regions which is unsuitable in the current deregulated electricity markets.

As in [12], the centralized problem (4.1) can be written in a hierarchical form as follows.

$$\begin{aligned} \min_{q} \quad & C_1\left(g_1^*(q)\right) + C_2\left(g_2^*(q)\right) \\ \text{subject to} \quad & q \le Q \qquad\qquad (\mu_q) \end{aligned} \tag{4.2}$$

where $g_n^*(q)$, $n \in \{1, 2\}$, is the optimal dispatch for area n, given the interchange level q.

The regional dispatch problem for area 1 is specified as

$$\begin{aligned} \min_{g_1 \in \mathcal{G}_1} \quad & c_1(g_1) \\ \text{subject to} \quad & \mathbf{1}^\intercal(d_1 - g_1) + q = 0 \qquad (\lambda_1) \\ & S_1(d_1 - g_1) + s_1 q \leq F_1 \quad (\mu_1) \end{aligned} \tag{4.3}$$

and for area 2 as

$$\begin{aligned} \min_{g_2 \in \mathcal{G}_2} \quad & c_2(g_2) \\ \text{subject to} \quad & \mathbf{1}^\intercal(d_2 - g_2) - q = 0 \qquad (\lambda_2) \\ & S_2(d_2 - g_2) - s_2 q \leq F_2. \quad (\mu_2) \end{aligned} \tag{4.4}$$

Note that this optimization involves an outer problem (4.2) to optimize the interchange level q, and an inner problem that is naturally decomposed into two regional problems, both parameterized by q. In other words, the optimizer and associated Lagrangian multipliers for (4.3) and (4.4) are functions of q, *i.e.*, $g_n^*(q)$, $\lambda_n^*(q)$, $\mu_n^*(q)$, $n \in \{1, 2\}$.

4.3.2 Tie Optimization

The key idea of TO [47] is to determine the interchange schedule by intersecting the demand and supply curves of interchange. By the demand/supply curve we mean the incremental cost of generation for each control area at different interchange levels. Each point on the demand/supply curve is essentially the LMP at the proxy bus at the given interchange level.

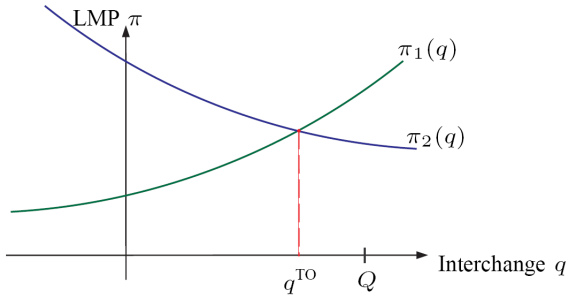


Figure 4.4: Optimal schedule q^{TO} under TO scheme.

Mathematically, given the interchange level q, the LMP at the proxy bus for area n is defined as

$$\pi_n(q) = \lambda_n^*(q) + s_n^{\mathrm{T}}\mu_n^*(q), \tag{4.5}$$

where $\lambda_n^*(q)$ and $\mu_n^*(q)$, $n \in \{1, 2\}$, are the Lagrangian multipliers associated with the optimal solution of (4.3) and (4.4). Since $\pi_n(q)$ indicates area n's incremental cost to dispatch up or down around q, we can view it as a supply curve for the exporting area or a demand curve for the importing area[2].

We use the graphical representation in [47] to illustrate the basic principle of TO. As shown in Figure 4.4, curve $\pi_n(q)$ represents the incremental cost of generation for region n and Q is the interface capacity. In this example, the direction of interface flow[3] is from area 1 to area 2, so $\pi_1(q)$ and $\pi_2(q)$ serve as supply and demand functions respectively. The optimal schedule q^{TO} is set at the intersection of the two curves. Note that if this quantity exceeds the interface capacity, the schedule should be set at the limit Q instead. The interface capacity constraint, in this case, becomes binding and price separation happens between markets. It should also be noted that import and export transactions are settled

[2]We note that $\pi_n(q)$ should be called the inverse supply/demand function using the standard economic terminology. Since this work does not involve any supply/demand function in the form of $q(\pi)$, we drop the term "inverse" without causing any confusion.

[3]The direction of interface flow can be determined by comparing the prices at $q = 0$: if $\pi_1(0) < \pi_2(0)$, the economically correct direction is from region 1 to 2; otherwise, the opposite.

at the *real-time LMPs* which are calculated at the proxy buses after the power delivery.

According to [47], the interchange schedule of TO is the optimal solution to (4.1) (as well as (4.2)). This intuitive argument is a manifestation of a deeper connection between social welfare optimization illustrated in Figure 4.4 and cost minimization defined by (4.1). In what follows, we will exploit this connection in the presence of uncertainty.

4.4 Stochastic Single Interface Scheduling

In order to incorporate system uncertainty, we propose a stochastic interchange scheduling technique, called Stochastic Tie Optimization (STO), based on the design of TO. The scheduling problem is formulated as a two-stage stochastic optimization that minimizes the expected overall operating cost. To develop a solution approach, we present an equivalent social welfare maximization from which the optimal interchange is obtained with the consideration of system uncertainty.

4.4.1 Stochastic Programming Formulation

Because the interchange is determined prior to generations and demands being realized, the interchange scheduling in the presence of generation and demand

uncertainty is fundamentally a two-stage stochastic optimization problem. The first stage involves minimizing the expected overall generation cost whereas, in the second stage, the regional optimal dispatch problems are solved given the interchange level q and the realizations of random net load d_1 and d_2, as defined in (4.3-4.4).

The two-stage stochastic optimization can be formulated as

$$\begin{aligned} \min_{q} \quad & \mathbb{E}_{d_1,d_2}\left[C_1\left(g_1^*(q,d_1)\right) + C_2\left(g_2^*(q,d_2)\right)\right] \\ & q \le Q \qquad\qquad (\mu_q) \end{aligned} \tag{4.6}$$

where the expectation is taking over all randomness of internal net loads $d_1 \sim \mathcal{F}_1$ and $d_2 \sim \mathcal{F}_2$. The optimal regional dispatch and the associated Lagrangian multipliers of (4.3) and (4.4) are now parameterized by two factors: the interchange level q and the net load realization d_n. So the LMP $\pi_n(q, d_n)$ at the proxy bus is a function of both q and d_n.

In general, problem (4.6) is intractable using standard optimization techniques when d_1 and d_2 have continuous distributions. In the following, we propose an indirect approach by solving a stochastic social welfare maximization using the same technique as TO.

4.4.2 Social Welfare Optimization

We now present an optimization problem from the import-export perspective, but taking into account that both import and export regions must agree on the

forward interchange quantity in the presence of future demand and supply uncertainty. Because the interchange quantity is fixed ahead of the actual power delivery, each region may have to rely on its internal resources to compensate uncertainty in real time. To this end, it is reasonable for the export region to maximize its *expected producer surplus* and the import region to maximize its *expected consumer surplus*.

Without loss of generality, let region 1 be the exporter. For fixed interchange q and demand d_n, let $\pi_n(q, d_n)$ be the LMP at the proxy bus for area n. The associated interchange social welfare is defined as the sum of the importer and the exporter surpluses, which is given by

$$\mathrm{SW}(q, d_1, d_2) \triangleq \int_0^q [\pi_2(x, d_2) - \pi_1(x, d_1)]dx. \tag{4.7}$$

Note that, at the time of determining the interchange q, random generations and demands in both regions are not yet realized. Therefore $\pi_1(q, d_1)$ and $\pi_2(q, d_2)$ are also random. We thus face a *stochastic optimization problem* in which the interchange q is a decision variable that can be set to maximize the *expected social welfare* $\mathbb{E}_{d_1,d_2}[\mathrm{SW}(q, d_1, d_2)]$.

The optimal interchange can be obtained by maximizing the expected interchange social welfare[4]

$$\max_{q \le Q} \int_0^q [\bar{\pi}_2(x) - \bar{\pi}_1(x)]\, dx \tag{4.8}$$

where $\bar{\pi}_n(q) \triangleq \mathbb{E}_{d_n}[\pi_n(q, d_n)]$ is the expected LMP—a function of the net interchange q—at the proxy bus of region i. The ways to compute or estimate these expected functions are presented in Section 4.4.4.

[4]Note that the expectation and integration in $\mathbb{E}_{d_1,d_2}[\mathrm{SW}(q, d_1, d_2)]$, *i.e.*, the objective function

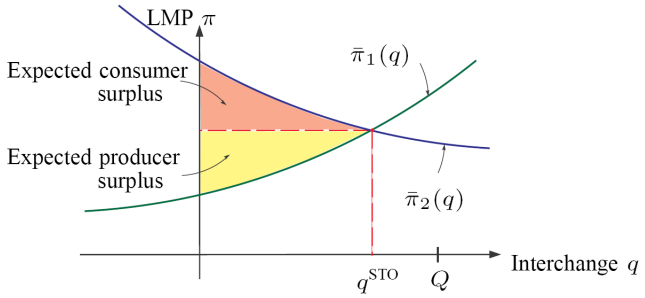


Figure 4.5: Optimal schedule q^{STO} under STO scheme.

We see from Figure 4.5 that problem (4.8) is the same as TO in Figure 4.4 with deterministic demand and supply functions replaced by their expectations. Therefore (4.6) can be solved easily using the same technique as in TO once expected demand and supply functions are given. This means that the optimal interchange can be determined by first searching for the intersection of $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$ and then checking whether the interface capacity constraint is satisfied.

4.4.3 Stochastic Tie Optimization

Although the equivalence of cost minimization and economic surplus maximization is intuitive in the deterministic setting, it is not obvious that the equivalence of the two stochastic programs holds. In this section, we establish the equivalence of (4.6) and (4.8).

Theorem 6. *If problem (4.3) and (4.4) are not degenerate for all d_1, d_2, and $q \leq Q$, then problems (4.6) and (4.8) have the same optimizer q^* satisfying*

$$\bar{\pi}_1(q) = \bar{\pi}_2(q)$$

if q^* and Q otherwise.

in (4.8), is interchangeable.

Proof. See Appendix A.3.1. □

Theorem 6 provides a new solution approach to the intractable problem (4.6). Instead of solving (4.6) directly, we only need to compute (forecast) the expected price curves and find the intersection using a searching algorithm such as the binary search. Since these (expected) price functions do not include the confidential information (bids and offers) from each region, the interchange problem can be solved by one of the operators as long as the other operator shares its expected price curve. In such a process, operators do not need to iteratively update or exchange information during the scheduling process. This property, inherent from TO, is in contrast to most decomposition methods where subproblems are resolved and intermediate results are exchanged in each iteration. Because one-time information exchange is sufficient for the optimal schedule, operators do not need to repeatedly solve regional OPF problems which can be computationally expensive for large systems. Such a property significantly reduces the real-time computational efforts, thereby providing the potential of higher scheduling frequency and alleviating the latency risk caused by system uncertainty.

4.4.4 Computation of Expected LMP

Here we discuss three techniques to compute or estimate the expected price curve.

First, the expected demand and supply curves can be computed using the

probabilistic forecasting technique presented in Chapter 3. Specifically, the expected LMP $\bar{\pi}_n(q)$ at the proxy bus in area n can be computed from the distribution of net load d_n at each level of interchange q. Such an approach, although providing exact expression of the expected demand and supply functions, may be computationally costly due to the fact that there may be a large number of critical regions. In practice, one may only need to compute a small subset of critical regions that contain typical net load realizations, *e.g.*, around the mean value of the net load.

A second approach is to obtain expected demand and supply curves based on an interpolation of sampled expected demands and supplies. In general, the expected price function $\bar{\pi}_n(q)$ may not have analytical forms; a set of interchange-price pairs is sufficient for the scheduling purpose. For example, operators may only compute the expected LMPs, $\{\bar{\pi}_1^k\}_{k=1}^K$ and $\{\bar{\pi}_2^k\}_{k=1}^K$, for a finite number of interchange levels $\{q^k\}_{k=1}^K$. To determine the intersection, the operator of region n shares the set of pairs $\{(\bar{\pi}_n^k, q^k)\}_{k=1}^K$ to the other operator who can then use the interpolation technique to approximate the entire curve $\bar{\pi}_n(q)$.

Yet another approach is based on the use of historical data. Currently, external transaction data is not only accessible to the operators but also to all market participants. For example, the external transaction data associated with the CTS interface between New York ISO and ISO New England, is publicly available: New York ISO provides the external CTS price online[41] and ISO New England updates the real-time scheduled interchange on the ISO's website[32]. Given the enormous amount of historical data, regression models can be used as a reasonable estimate of the relationship between interchange quantities and expected

prices, which can be treated as the expected supply and demand functions.

4.4.5 Summary of Scheduling Procedure

We conclude this section by describing the procedure of the proposed scheduling technique. Without loss of generality, we assume that the operator of area 1 (ISO 1) submits its price curve to the operator of area 2 (ISO 2) who then determines the interchange schedule.

The scheduling procedure of STO is summarized as follows:

(1) operators compute the expected price functions $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$;

(2) ISO 1 submits $\bar{\pi}_1(q)$ to ISO 2;

(3) ISO 2 determines the direction of the interface flow, computes q^{STO} as defined in Theorem 6, and communicates the schedule to ISO 1;

(4) both operators dispatch their internal resources to implement the agreed interchange level in real time.

We point out that the expected price curve $\bar{\pi}_n(q)$ produced in Step (1) covers both positive and negative interchange values. The direction of the interface flow is determined by the sign of the cleared interchange. We note that the direction and volume of the optimal interchange can be obtained simultaneously from the intersection of the two expected curves which are defined on q with the same direction. For example, the direction of q is defined from area 1 to area 2

in (4.1). If the intersection of $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$ were set at a positive value $q_{\text{int}} > 0$, the interface flow direction would be from region 1 to region 2. Otherwise, a negative value $q_{\text{int}} < 0$ would be the opposite direction.

After agreeing the interchange schedule, operators will prepare their internal resources for the actual delivery. Note that the interchange schedule is determined in advance of net load realization. This time difference is approximately 15 minutes in [47] and up to 75 minutes in the current implementation of CTS. When load and renewable generations are realized in real time, operators will dispatch the internal resources to meet the regional demand and the agreed schedule.

4.5 Generalizations

In this section, we discuss three possible generalizations of the proposed STO scheduling technique. The first one is incorporating external market participants in the interchange coordination as the CTS proposal. The second is the multi-area (more than two areas) interchange scheduling. The last one is the consideration of regional reserve market in the multi-area system setting.

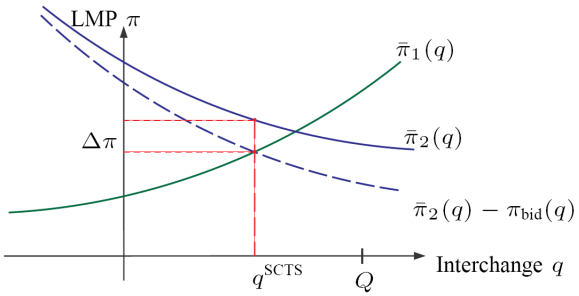


Figure 4.6: Optimal schedule q^*_{SCTS} under SCTS scheme.

4.5.1 Stochastic Coordinated Transaction Scheduling

In order to incorporate external market participants, we generalizes the CTS proposal currently in implementation to Stochastic Coordinated Transaction Scheduling (SCTS) using the same idea of STO. As in CTS, market participants are allowed to submit requests to buy and sell power simultaneously on both sides of the interface. Such a request is called an *interface bid*, which includes a price indicating the minimum expected price difference between two regions that the participant is willing to accept, a transaction quantity of the interchange, and a flow direction.

The graphical representation of SCTS is given in Figure 4.6. Analogous to Figure 4.5, $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$ are the expected LMP curves of region 1 and region 2, respectively. The third price curve[5] $\bar{\pi}_2(q) - \pi_{\text{bid}}(q)$ is the adjusted curve of $\bar{\pi}_2(q)$ by subtracting the aggregated interface bids $\pi_{\text{bid}}(q)$. The SCTS schedule is set at the intersection of $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q) - \pi_{\text{bid}}(q)$ in the absence of interface limit. All interface bids with the expected price difference less than $\Delta\pi = \pi_{\text{bid}}\left(q^{\text{SCTS}}\right)$

[5]To incorporate the aggregated interface bid $\pi_{\text{bid}}(q)$, one can subtract it from the expected demand curve $\bar{\pi}_2(q)$ as shown in Figure 4.6 or add it to the expected supply curve $\bar{\pi}_1(q)$. In either way, the optimal schedule is set at the intersection of the adjusted curve and the original curve of the other.

are accepted and settled at the real-time LMP difference. Mathematically, this problem is formulated as:

$$\max_{q \le Q} \int_0^q [\bar{\pi}_2(x) - \pi_{\text{bid}}(x) - \bar{\pi}_1(x)]dx \tag{4.9}$$

where the expected price functions $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$ are derived from (4.3) and (4.4). Note that problem (4.9) is not equivalent to the cost minimization (4.2) due to the introduction of market participants.

Compared to the original design of CTS, the only difference of SCTS is the use of the expected supply and demand curves. This implies that one-time information exchange is sufficient; no iteration between operators is necessary during the scheduling procedure.

4.5.2 Extension to Multi-Area Interchange Scheduling

Having considered the problem of interchange scheduling between a pair of neighboring ISOs, it is natural to examine possible extensions to three or more interconnected areas. Such a problem clearly has significant implications in practice. For instance, New York ISO has interfaces with both ISO New England and PJM, and PJM has interfaces with New York ISO and Midcontinent ISO. Thus effective coordination among all operators can affect the overall efficiency of a large-scale interconnected system.

The proposed scheduling technique is directly applicable to the *asynchronous multi-area interchange scheduling problem*. In particular, because operators run

their regional dispatches asynchronously, an ISO with multiple interfaces may determine interchange schedules one at a time. For such cases, the problem is effectively reduced to a *pairwise* interchange scheduling problem for which the proposed STO can be applied directly.

It is perhaps computationally more efficient, possibly with better performance, if multiple interfaces can be scheduled *simultaneously*. The problem of jointly optimizing all interface flows presents a nontrivial difficulty in extending the approach presented in this section, both conceptually and algorithmically. The key issue arises from the fact that the idea of intersecting the supply and demand curves is no longer applicable. In a way, the operator has to deal with high dimensional "supply and demand surfaces" on which one has to find the optimal *interchange vector*. It is not clear such a vector has the correct economic interpretation as in the one dimensional case. Even if it does, it is not obvious how to obtain such a multi-dimensional schedule in a computationally tractable fashion. In section 4.6, we propose an approach for stochastic interchange scheduling in the multi-area multi-interface system based upon the idea presented in this section.

4.5.3 Multi-Area Energy-Reserve Market

The proposed interchange scheduling technique assumes that each area maintains its own reserve requirement and reserve allocation rule based on each region's operating criteria. In this way, the scheduling process will require minimum or no change to the current industry practice. Maintaining separate re-

serve models (or reserve markets) in each area does not affect the proposed interchange scheduling process, but the average supply and demand curves can be different from those obtained without reserve requirements.

Here we describe the general idea how the proposed interchange scheduling technique incorporates regional reserve markets. The multi-area joint energy and reserve market clearing problem under uncertainty can be still formulated as a two-stage stochastic program: in the first stage, the interchange is optimized for the expected overall cost; and for the second stage, instead of optimizing the energy dispatch, the reserve levels are also determined via *an energy and reserve co-optimization*. For energy and reserve co-optimization model, refer to [49]. The scheduling procedure is exactly the same as the summary given in Section 4.4.5, except that the expected price is computed from the energy and reserve co-optimization model instead of energy-only optimization.

4.6 Stochastic Multi-Interface Scheduling

In this section, we generalize the single interface scheduling algorithm to the multi-area system setting where multiple interface flows are involved, as illustrated in Figure 4.3. In particular, the proposed interface-by-interface scheduling (IBIS) algorithm is specialized for synchronous scheduling and asynchronous scheduling.

4.6.1 Synchronous Interchange Scheduling

The first scenario we consider is the synchronous interchange scheduling in which all operators in the interconnected system have a unified timetable, *i.e.*, all interface flows are optimized simultaneously at each scheduling time.

Consider an interconnected system with N independently operated areas (of an arbitrary network topology) and I separate scheduling interfaces. The multi-area synchronous interchange scheduling problem, analogous to the single interface scheduling problem, is also a two-stage stochastic optimization: the first stage is to set the values of all interface flows by minimizing the expected overall cost; the second stage is to minimize the cost of individual areas given the fixed interchange and realized random generation and demand.

As a generalization of (4.6) for the single interface scheduling problem, the first stage optimization for the multi-area system is given by

$$\begin{aligned} \min_{q} \quad & \bar{C}(q) = \textstyle\sum_{n=1}^{N} \mathbb{E}_{d_n} \left[C_n \left(g_n^*(q_n, d_n) \right) \right] \\ & q \le Q \qquad\qquad (\mu_q) \end{aligned} \tag{4.10}$$

where the interchange vector[6] q is a real vector in dimension I, q_n the vector of interface flows associated with area n assuming *outbound* directions, $\bar{C}(q)$ the expected overall system cost, and $g_n^*(q_n, d_n)$ the optimal regional dispatch in area n, given the interchange level q_n over the interfaces associated with area n and the realized net load d_n.

[6]Note that the interchange vector q in (4.10) includes all interface flows with predetermined directions. This is the generalization of the scalar case in formulation (4.6) where the interchange vector q, representing the interface flow from area 1 to 2, is in dimension 1.

In the second stage, each operator dispatches the internal resource to meet the interchange schedule q_n and the realized internal net load d_n in the least cost manner subject to the generation and transmission constraints. The optimization problem for area n, $n = 1, 2, \cdots, N$, is specified as

$$\begin{aligned} \min_{g_n} \quad & C_n(g_n) \\ \text{subject to} \quad & \mathbf{1}^\intercal(d_n - g_n) + \mathbf{1}^\intercal q_n = 0 \quad (\lambda_n) \\ & S_n(d_n - g_n) + \tilde{S}_n q_n \leq F_n \quad (\mu_n) \\ & g_n \in \mathcal{G}_n \end{aligned} \tag{4.11}$$

where $\tilde{S}_n$ is the shift factors of proxy buses associated with area n to internal transmission lines in area n

Given the first stage decision q_n and the realization of net load d_n, the second stage problems are naturally decoupled and can be solved by their own operators. The LMP vector π_n at the proxy buses for area n is calculated from the Lagrangian multipliers of (4.11)

$$\pi_n(q_n, d_n) = \mathbf{1}\lambda_n^*(q_n, d_n) + \tilde{S}_n^\intercal \mu_n^*(q_n, d_n) \tag{4.12}$$

where $\lambda_n^*(q_n, d_n)$ and $\mu_n^*(q_n, d_n)$ are functions of the realization d_n and the first stage decision q_n.

The expected multi-dimensional LMP function $\bar{\pi}_n(q_n)$ for area n is therefore defined as

$$\bar{\pi}_n(q_n) = \mathbb{E}_{d_n}[\mathbf{1}\lambda_n^*(q_n, d_n) + \tilde{S}_n^\intercal \mu_n^*(q_n, d_n)] \tag{4.13}$$

where the expectation is taking over all randomness of the net load $d_n \sim \mathcal{F}_n$.

Interface-by-Interface Scheduling

The idea of the proposed scheduling algorithm is to iteratively optimize the interchange vector, one interface at a time, until the termination criterion satisfied. Let $q^{(k)}$ be the vector of all interface flows at iteration k and $q^{(k)}(i)$ the ith element of $q^{(k)}$ representing the flow over the ith interface. The optimization for the ith interface flow at iteration k is given by

$$q^{(k)}(i) = \underset{q(i)\leq Q(i)}{\arg\min}\ \bar{C}\left(q(i), q^{(k)}(-i)\right) \tag{4.14}$$

where $Q(i)$ is the transmission capacity for interface i and $q^{(k)}(-i)$ the vector of $q^{(k)}$ with the most updated values after removing the ith entry, *i.e.*,

$$q^{(k)}(-i) \triangleq (q^{(k)}(1), \cdots, q^{(k)}(i-1), q^{(k-1)}(i+1), q^{(k-1)}(I)). \tag{4.15}$$

By Theorem 6, the optimal solution $q^{(k)}(i)$ can be obtained by searching the intersection of the expected supply and demand function defined in (4.13) for $q(i)$ with fixed $q^{(k)}(-i)$ and check if the interface capacity $Q(i)$ is satisfied.

Formally, the synchronous interface-by-interface scheduling (SIBIS) algorithm to obtain the optimal interchange to (4.10) is given in Algorithm 2.

In practice, a positive value is chosen for ϵ to ensure a finite termination of SIBIS. When ϵ is set to zero, the optimality and convergence behavior of SIBIS can be proved (the proof of Theorem 7 is given in Appendix A.3.3).

Theorem 7. *Let $\{q^{(k)}\}_{k=0}^{\infty}$ be the sequence generated by Algorithm 2 with $\epsilon = 0$. Then, every limit point of $\{q^{(k)}\}_{k=0}^{\infty}$ is optimal to (4.10).*

Algorithm 2 Synchronous Interface-by-Interface Scheduling

1: **given** a feasible initial point $q^{(0)}$, the expected LMP function $\bar{\pi}_n(q_n)$ for area n, $n = 1, \cdots, N$, and a tolerance $\epsilon \geq 0$.
2: **repeat**
3: $k = k + 1$, $q^{(k)} = q^{(k-1)}$.
4: **for** $i = 1, 2, \cdots, I$ **do**
5: Obtain $q^{(k)}(i)$ in (4.14) by intersecting the expected supply and demand functions.
6: **end for**
7: **until** $\|q^{(k)} - q^{(k-1)}\|_2 \leq \epsilon$.

It should be noted that SIBIS is a form of cyclic coordinate descent method in which one cyclically iterates through the directions, one at a time, minimizing the objective function with respect to each coordinate direction at a time. The early study of the coordinate descent method dates back to 1950s [31]. The convergence of the method has been extensively studied in the literature [31, 38, 45] under various assumptions. Given the strict convex assumption of the regional cost function $C_n(g_n)$, the objective function $\bar{C}(q)$ is a continuously differentiable convex function, as shown in the proof of Theorem 7. If $\bar{C}(q)$ has local strict convexity in the feasible region of (4.10), linear rate of convergence can be established as the case in [38].

4.6.2 Asynchronous Interchange Scheduling

The second scenario we consider is the asynchronous scheduling in which an operator with multiple interfaces determines interface flow one at a time. For such cases, the multi-interface scheduling problem is effectively reduced to a sequential single interface flow optimization.

Let q^t be the vector of all interface flows at time t, $q^t(i)$ the ith element of q^t and $q^t(-i)$ the vector of q^t after removing $q^t(i)$. Given the interface flow vector $q^{t-1}(-i)$ at time $t-1$, the schedule of the ith interface $q^t(i)$ at time t is given by

$$q^{(t)}(i) = \underset{q(i)\leq Q(i)}{\arg\min} \sum_{n=1}^{N} \mathbb{E}_{d_n^t}\left[C_n\left(g_n^*\left(q(i), q^{t-1}(-i), d_n^t\right)\right)\right] \tag{4.16}$$

where the expectation is taking over the randomness of net load d_n^t with respect to the distribution $\mathcal{F}_n^t$ at time t, and the regional dispatch $g_n^*(q_n), d_n^t)$ for area n is the optimal solution to (4.11), given the interchange schedule q_n and the realization of net load d_n^t.

Since the interface flow $q(i)$ is the only decision, the objective function in (4.16) only involves two areas intersects at the ith interface. Therefore, the optimal interface flow $q^t(i)$ at time t can be obtained by intersecting the expected supply and demand function of $q(i)$ given the distribution $\mathcal{F}_n^t$ of the random net load d_n^t with the other interface flows fixed at $q^{t-1}(-i)$. Note that the expected LMP function $\bar{\pi}_n^t(q_n)$, similarly defined in (4.13), depends on time through the distribution $\mathcal{F}_n^t$.

The distinction between synchronous and asynchronous scheduling lies in the decision at each scheduling time. For synchronous scheduling, the entire interchange vector is optimized via the iterative process given in Algorithm 2 at each scheduling time t. For the asynchronous scheduling, on the other hand, only one element of the interchange vector is optimized at time t. Therefore, the solution of the asynchronous scheduling algorithm at time t is suboptimal in terms of minimizing the expected overall system cost.

The asynchronous interface-by-interface scheduling (AIBIS) algorithm is for-

Algorithm 3 Asynchronous Interface-by-Interface Scheduling

1: **given** a feasible initial point $q^{(0)}$, the expected LMP function $\bar{\pi}_n^t(q_n)$ for area n, $n = 1, \cdots, N$, at time t, and a termination time T.
2: **repeat**
3: $t = t + 1$, $i = t \bmod I$.
4: Obtain the optimal solution $q^t(i)$ of (4.16) by intersecting the expected supply and demand function.
5: **until** $t = T$.

mally described in Algorithm 3. It should be noted that the iterative process of AIBIS is carried out over time which should be distinguished with that in SIBIS.

We note that if the net load process d_n^t, $n \in \{1, \cdots, N\}$, is independent in time, the optimal interchange depends only on the *marginal distribution* of the random load at the time of delivery. If in addition, the process is stationary, *i.e.*, d_n^t is independent and identically distributed (i.i.d.), then the optimal interchange is constant. In this case, AIBIS is essentially the classical cyclic coordinate decent spread over time. In comparison with SIBIS, SIBIS achieves optimal interchange for at every time whereas AIBIS achieves the optimality over time. The following Theorem, whose proof is given in Appedix A.3.4, formalize this argument.

Theorem 8. *Let $\{q^t\}_{t=0}^{\infty}$ be the sequence generated by Algorithm 3 with $T = \infty$. If the net load $d_n^t \overset{i.i.d}{\sim} \mathcal{F}_n$ for all n, then every limit point of $\{q^t\}_{t=0}^{\infty}$ is optimal to (4.10).*

When d_n^t is not i.i.d., the interchange sequence generated by AIBIS does not converge to that by SIBIS. The performance (averaged over time) of AIBIS algorithm does not in general converge to that of SIBIS; the lack of synchronization translates to a performance loss. When the load process is a finite state Markov chain, however, a modification of AIBIS that separately adapts the interchange for different load state will have the same time averaged performance as that of

SIBIS.

4.7 Evaluation

In this section, we present numerical results of the proposed scheduling techniques using various testing systems with different topologies.

We first compare the performance of the proposed STO in Section 4.4 with that of TO on two systems: a 2-area 6-bus system and a 2-area 118-bus system. In both examples, TO uses the certainty equivalent forecast of wind power, *i.e.*, the mean value, while STO uses the probabilistic forecast, *i.e.*, the distribution.

The proposed algorithms for the multi-interface scheduling are then evaluated on the IEEE 3-area 118-bus system. The proposed SIBIS is compared with the certainty equivalence (CE) technique,[7] which uses the mean value of the random variable to schedule the interchange. Performance comparison between SIBIS and AIBIS is followed.

[7]The CE method also adopts the iterative procedure given in Algorithm 2. The only difference lies in the price functions to obtain the single interface flow. Instead of using the expected supply/demand function $\mathbb{E}_{d_n}[\pi_n(q_n, d_n)]$, the CE method uses the supply/demand function $\pi_n(q_n, \mathbb{E}_{d_n}[d_n])$ by substituting the random variable by its the expected value.

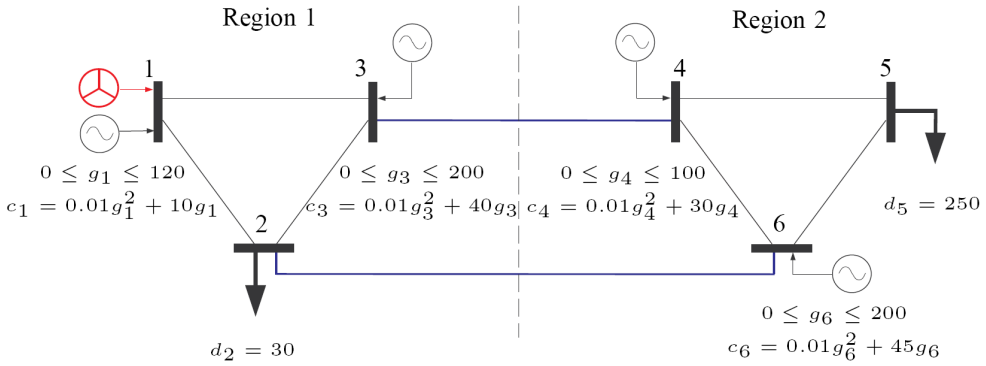


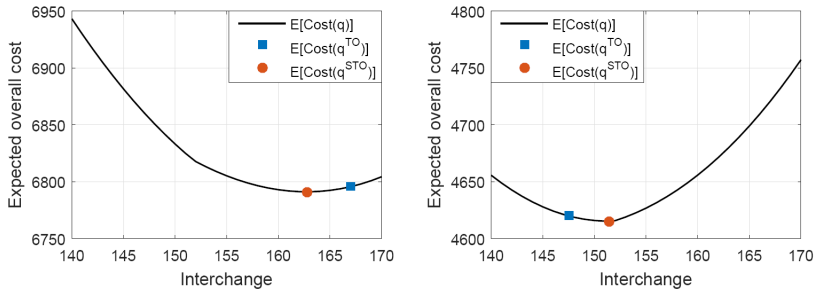
Figure 4.7: A 2-region 6-bus system.

4.7.1 Example 1: A 2-Area 6-Bus System

Consider a 2-area 6-bus system as depicted in Figure 4.7. Generator incremental cost functions, capacity limits, and load levels (the default values) are presented in the figure. All lines were assumed to be identical except for the maximum capacities: the tie lines (line 2-6 and line 3-4) and the internal transmission lines in region 1 (line 1-2, line 1-3, and line 2-3) had the maximum capacities of 100 MW, and the internal lines in region 2 have the maximum capacities of 200 MW. The system randomness came from the wind generator at bus 1 in region 1. The entire network model (the shift factor matrix) was assumed to be known to both operators. Bus 3 and 6 were chosen as proxy buses.

Baseline

We first tested a baseline with the probabilistic wind (forecast) distribution $\mathcal{N}(55, 10^2)$. Two load levels were chosen to illustrate the two inefficiency symptoms of TO schedule: the first load level $d_5 = 250$ was an example of the counter-



(a) $d_5 = 250$ (b) $d_5 = 200$

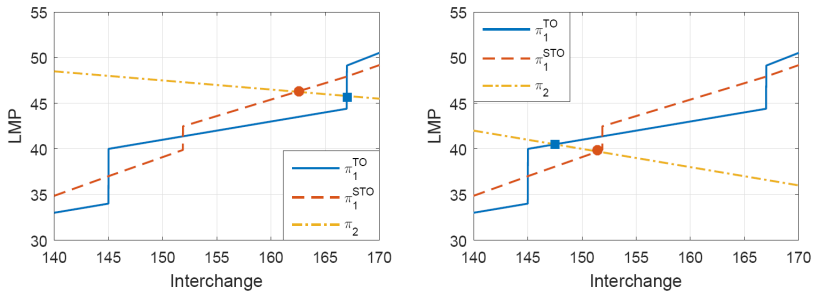
Figure 4.8: Expected overall cost curves with different load levels.

Table 4.1: Performance comparison of TO and STO

Scenario	Method	q	$\mathbb{E}[\text{Cost}(q)]$	$\mathbb{E}[\Delta\pi(q)]$
$d_5 = 250$	TO	167	6795.6	−2.3
	STO	162.8	6791.1	0
$d_5 = 200$	TO	147.5	4621.6	2.5
	STO	151.4	4608.6	0

intuitive flow, and the second load level $d_5 = 200$ showed the case of interface under-utilization. All results were generated from Monte Carlo simulation. The actual wind realizations were sampled from the forecast distribution $\mathcal{N}(55, 10^2)$ as well.

We plotted the expected overall operating costs at different interchange levels in Figure 4.8. At the interchange schedule of STO (the red circle), the expected overall system cost was minimized in both cases. This verified the optimality of the proposed scheduling technique. Table 4.1 provides more detailed statistics for the schedule of TO and that of STO. Under STO, the expected prices at the two proxy buses converged in both cases. Under TO, on the other hand, the price disparity occurred in both cases. In the first example, the expected price difference was −2.3 \$/MWh, which means the expected price of the importing region (region 2) was lower than that of the exporting region. This im-



(a) $d_5 = 250$ (b) $d_5 = 200$

Figure 4.9: Expected price curves under different load levels.

plies that the interchange was scheduled from a high cost region to a low cost region, which is economically counter-intuitive. In the second example, the expected price difference at the interchange level of TO schedule was 2.5 \$/MWh, *i.e.*, the marginal price of the importing region was higher than that of the exporting region. With this price difference, increasing the interchange level can further reduce the expected overall cost, meaning that the interchange capacity was underutilized. In fact, because of the optimality of STO schedule, any schedule higher than this optimal level would cause counter-intuitive flows, and any schedule lower than that would lead to the interface under-utilization.

To see why these two phenomena happen, we took a closer look at the supply and demand curves given in Figure 4.9 where π_1^{TO} and π_1^{STO} are the supply curves of region 1 for TO and STO respectively. Since there was no randomness in region 2, π_2 is the deterministic demand curve for both TO and STO. If there were no uncertainty of the wind forecast, *i.e.*, the variance of wind generation distribution were zero, the expected supply curve would be exactly the same as π_1^{TO}. So the increase of the standard deviation of wind generation changes the expected supply curve from π_1^{TO} to π_1^{STO}.

Figure 4.9 shows some insights into the inefficiency associated with the cer-

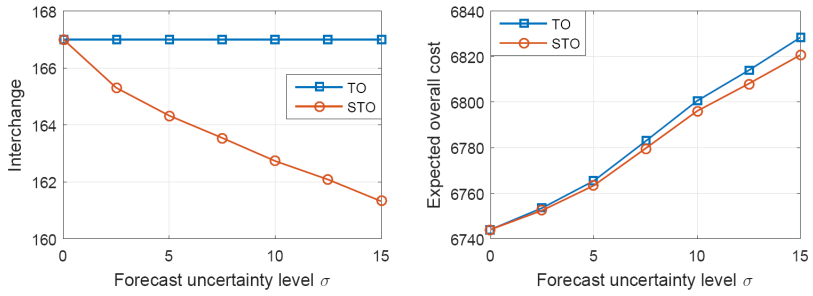
tainty equivalent approach of TO. Here we explain the case of $d_5 = 250$ in Figure 4.9a with details. TO sets the interchange by intersecting the supply curve (solid) from region 1 with the demand curve (dash-dot) from region 2. This solution would have been exactly the same as STO if the wind production has zero variance.

In the presence of uncertainty, the expected supply curve (dash) in Figure 4.9a represents a higher cost of generation from region 1 around the interchange quantity set by TO. This means that, the 167 megawatt flow from region 1 to 2 set by TO is a flow from the higher cost region to the lower cost region, which is a case of counter-intuitive flow. STO, on the other hand, sets the interchange at a lower level (red circle) and there is no price differential at the proxy buses.

Similar rationale is also applicable for the case of $d_5 = 200$ in Figure 4.9b. At the TO interchange (the blue square), the expected generation cost of region 1 is lower than that of region 2. This means that the overall cost can be further reduced if more power flow from region 1 to region 2.

Impact of Forecast Uncertainty

The impact of the forecast uncertainty level was investigated by varying the standard deviation σ of the wind production distribution $\mathcal{N}(55, \sigma^2)$. Loads were set at the default values given in Figure 4.7.



(a) Interchange (b) Expected overall cost

Figure 4.10: Impact of the forecast uncertainty σ.

As shown in Figure 4.10, the interchange schedule under TO did not change with σ because it only used the mean value 55 of the wind production forecast. STO, on the other hand, captured the uncertainty level of the probabilistic forecast and adjusted the interchange schedule accordingly. The expected overall cost increased with the forecast uncertainty, which was observed in both TO and STO. Such trend is intuitive because uncertainty creates more inefficiencies. When there were no uncertainty ($\sigma = 0$), the schedules of TO and STO would be exactly the same.

4.7.2 Example 2: A 2-Area 118-Bus System

We divided the standard IEEE 118-bus system[8] into two areas: area 1 includes bus 1-12 (and associated internal transmission lines) and area 2 bus 13-118 (and associated internal transmission lines). In this system, there was a single interface consisting of five tie lines. Generator incremental cost functions, capacity limits, and load levels were set at the default values given in "case118" in [51]. Bus 4 and 69 were selected as the proxy buses. We imposed the maximum ca-

[8]All bus and branch indices are referred to [51].

pacity of 100 MW on line 8, 126 and 155. The interface transmission was not limited by default, but the impact of the interface congestion would be studied.

We introduced ten wind generators[9] in the two-region system, one in area 1 and nine in area 2. The wind power productions in each region were assumed to be independent random variables with identical probability distribution. We used a generalization of scenario-based model. In particular, suppose that there is a set of wind generation scenarios, each associated with a Gaussian model with mean μ_k and variance σ_k^2. If scenario k occurs with relative frequency p_k, then the distribution of the wind generation in region i is given by the so-called Gaussian mixture distribution

$$w_i \sim \sum_k p_k \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(w_i-\mu_k)^2}{2\sigma_k^2}}. \tag{4.17}$$

In this example, we only considered two scenarios: high wind scenario (characterized by the Gaussian distribution with mean μ_i^{high} and the standard deviation σ_i^{high} for region i) and low wind scenario (characterized by the Gaussian distribution with mean μ_i^{low} and standard deviation σ_i^{low} for region i). We denote the probability of high wind scenario by p and that of low wind scenario by $1 - p$. Therefore, the mean production value for each generator in region i is:

$$\bar{w}_i = p\mu_i^{\text{high}} + (1-p)\mu_i^{\text{low}}. \tag{4.18}$$

For the following simulation results, we used the continuous distribution (4.17) to compute the interchange of STO, and the mean value $\bar{w}_i$ for TO. All

[9]The wind generators were assumed to be located at bus 6, 90, 100, 103, 104, 105, 107, 110, 111, and 112.

Table 4.2: Wind distribution parameter value

p	Region	High wind	Low wind
0.5	w_1	$\mathcal{N}\left(350, 35^2\right)$	$\mathcal{N}\left(50, 5^2\right)$
	w_2	$\mathcal{N}\left(150, 15^2\right)$	$\mathcal{N}\left(50, 5^2\right)$

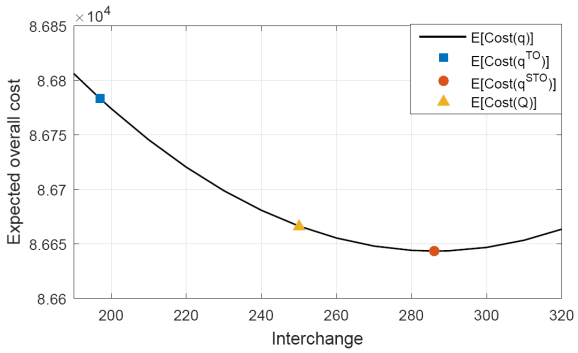


Figure 4.11: Expected overall cost: impact of interface congestion.

realizations of wind generation were sampled from the continuous distribution (4.17) using parameter values in Table 4.2.

Impact of Interface Congestion

In this case, we first verified the optimality of STO schedule in the absence of interface capacity and then examined the effect of interface congestion.

We plotted the expected overall operating costs at different interchange levels in Figure 4.11. The result verified the optimality of STO and the inefficiency of TO: the interchange level of STO was located at the minimum of the expected cost curve while the schedule of TO resided at a suboptimal point. The root cause of the inefficiency of TO, in this case, was the under-utilization of the interface transfer capacity.

Knowing that under-utilization causes inefficiency, an interesting question to ask is what happens if the interface capacity is limited and the optimal schedule of STO is not achievable. To answer this question, we imposed the limit of external injection/withdraw at the proxy bus to be 250 MW. From Figure 4.11, the expected cost at this limit (yellow triangle) was larger than that of the original STO schedule (red circle), but still better than that of TO (blue square). One can imagine: if the interface capacity were further restricted, for example, at a level smaller than the schedule of TO, both TO and STO would yield the same schedule that equals to this tight limit. The expected cost would increase for sure due to this restriction.

To scrutinize the impact of interface congestion, we calculated the expected price differences between two regions at different interchange schedules. Results are presented in Table 4.3. In the absence of interface constraint ($Q = \infty$), the LMPs at proxy buses converged at the schedule of STO, which created an ideal seamless market. However, when the interface was congested ($Q = 250$), the price disparity occurred: the exporting price was lower than the importing price by 1.27 \$/MWh. Economically, increasing the interface flow would reduce the price difference and improve the overall cost in expectation. Technically, the interface flow was constrained by its physical limit ($Q = 250$), so the schedule was stuck at this suboptimal point. Therefore, whenever the interface is congested, the price disparity occurs and the congestion cost increases.

Table 4.3: Impact of interface congestion

Q	Method	q	$\mathbb{E}[\text{Cost}(q)]$	$\mathbb{E}[\Delta\pi(q)]$
∞	TO	197.11	86782.72	3.14
	STO	286.11	86643.26	0
250	TO	197.11	86782.72	3.14
	STO	250	86666.14	1.27

Table 4.4: Interchange schedules and expected overall costs under different proxy buses

Proxy Bus[10]	q		$\mathbb{E}[\text{Cost}(q)]$	
	TO	STO	TO	STO
(4, 69)	197.11	286.11	86782.72	86643.26
(6, 24)	192.08	286.28	86873.96	86719.66
(10, 49)	196.88	285.65	87123.66	86985.48
(8, 30)	216.69	283.84	86636.51	86552.82
(11, 13)	64.79	199.79	87743.53	87247.55
(12, 14)	70.01	210.68	87634.56	87152.42

Impact of Proxy Bus Location

As pointed out in [28], all existing LMP based markets utilize proxy bus representation for the interchange modelling and scheduling. The location of proxy bus, however, is selected according to the industrial experience and there does not exist a common selection criterion. Here we tested various locations to investigate the impact of proxy bus locations. The distribution given in (4.17) with parameter values given in Table 4.2 was used.

From results presented in Table 4.4, we observed that both techniques were

[10]The pair (p_1, p_2) indicates the proxy bus locations in region 1 and region 2 respectively. For example, the pair (4, 69) means that bus 4 was used to represent the network in region 1 and bus 69 the network in region 2.

sensitive to the proxy bus locations. The optimal schedule under TO was more random than that under STO. For example, the schedule of TO at (8, 30) was three times larger than that at (11, 13) and (12, 14).

In this section, we present numerical results of the proposed scheduling algorithms on the IEEE 118-bus system. The performance of the proposed SIBIS algorithm is compared with the certainty equivalence (CE) technique,[11] which uses the mean value of the random variable to schedule the interchange. The comparison of SIBIS and AIBIS algorithms is then presented for time varying random processes.

4.7.3 Example 3: IEEE 3-Area 118-Bus System

In this example, we tested the proposed algorithms, SIBIS and AIBIS, for the multi-interface scheduling on the IEEE 3-area 118-bus system. The topology and area partition of the IEEE 118-bus system are given in Figure 4.12. The interchange vector q of this 3-area 2-interface system includes two interface flows where $q(1)$ and $q(2)$ are the flows to area 2 from area 1 and area 3, respectively. The load profile, generator capacities, cost functions, and line and bus labels were defined in "case118" in [51]. We imposed the maximum capacity of 100 MW on transmission line 8, 126 and 155, and 1200 MW on the two interfaces. Bus 31, 66 and 92 were selected as proxy buses for area 1, 2, and 3 respectively.

[11] The CE method also adopts the iterative procedure given in Algorithm 2. The only difference lies in the price functions to obtain the single interface flow. Instead of using the expected supply/demand function $\mathbb{E}_{d_n}[\pi_n(q_n, d_n)]$, the CE method uses the supply/demand function $\pi_n(q_n, \mathbb{E}_{d_n}[d_n])$ by substituting the random variable by its the expected value.

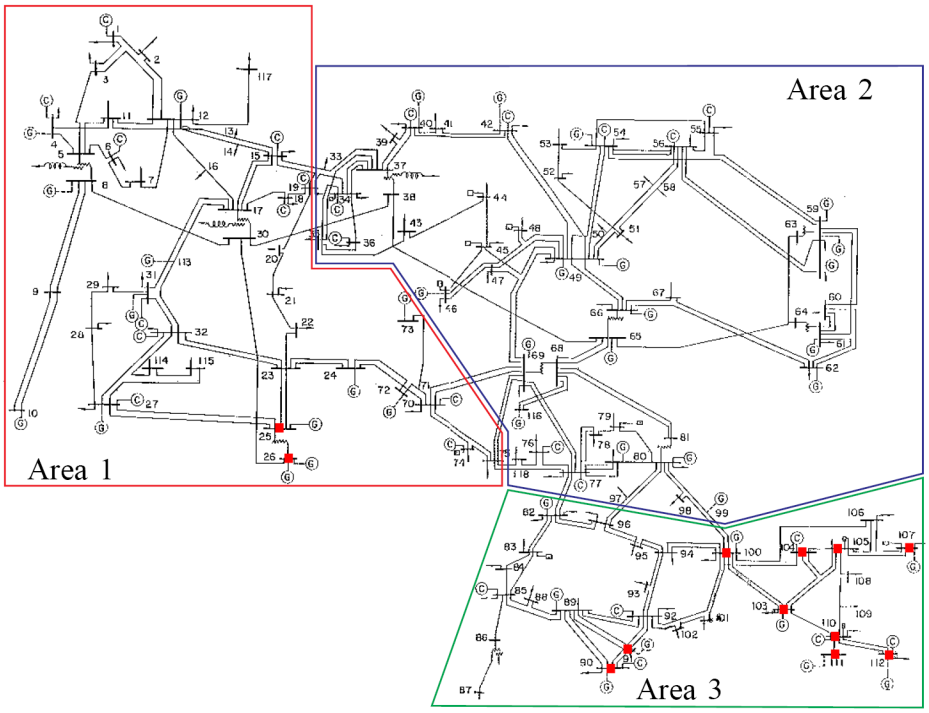


Figure 4.12: IEEE 3-area 118-bus system [7].

The system uncertainty arose from 12 wind generators (roughly 10% of buses) located at bus 25, 26, 90, 91, 100, 103, 104, 105, 107, 110, 111, and 112, as indicated by red boxes in 4.12. The selection of these locations were intended to simulate two wind farms; the small one has 2 wind generators in area 1, and the large one has 10 wind generators in area 3 concentrated on a few neighboring buses. All wind generators were assumed to be identical and follow a two-mode Gaussian mixture distribution whose probability density function is given by

$$f(w) = 0.5\frac{\exp\{-\frac{(w-\mu_h)^2}{2\sigma_h^2}\}}{\sqrt{2\pi\sigma_h^2}} + 0.5\frac{\exp\{-\frac{(w-\mu_l)^2}{2\sigma_l^2}\}}{\sqrt{2\pi\sigma_l^2}} \tag{4.19}$$

where the Gaussian distribution $\mathcal{N}(\mu_h, \sigma_h^2)$ represents the high wind scenario and $\mathcal{N}(\mu_l, \sigma_l^2)$ the low wind scenario.

Synchronous Interchange Scheduling

We first show the optimality and convergence behaviour of the proposed SIBIS and the two most common symptoms of inefficient schedule.

The distribution (4.19) was used as the probabilistic wind production forecast with parameter values $\mu_h = 150$, $\sigma_h = 12$, $\mu_l = 50$, and $\sigma_l = 4$. Note that the difference between the proposed algorithm and the benchmark technique is the use of wind production forecast. The synchronized scheduling algorithm uses the forecasted distribution while the certainty equivalence only uses the mean value $\bar{w} = 0.5\mu_h + 0.5\mu_l = 100$. The initial interchange vector was set at $q^{(0)} = (0, 1000)$ for both methods. Termination rule given in Algorithm 2 was used for both methods with tolerance $\epsilon = 0.001$.

Simulation 1: Tie Line Utilization

In this simulation, we examine the utilization of tie lines. Ideally, tie lie should be utilized fully to the extent that power flows from a lower price proxy to a high price proxy. All bus loads were set at the default values given in "case118" [51].

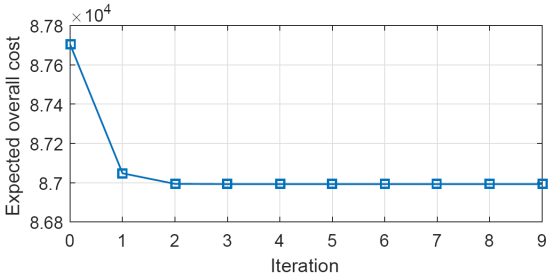
From the results shown in Table 4.5, we observed that the interface flows scheduled by CE and SIBIS had the same directions but different volumes. The CE schedule of transferring 703.32 MW from area 3 to area 2 resulted in the expected price disparity where the exporting area was \$31.33 while the import-

Table 4.5: Performance comparison in the interface under-utilization scenario

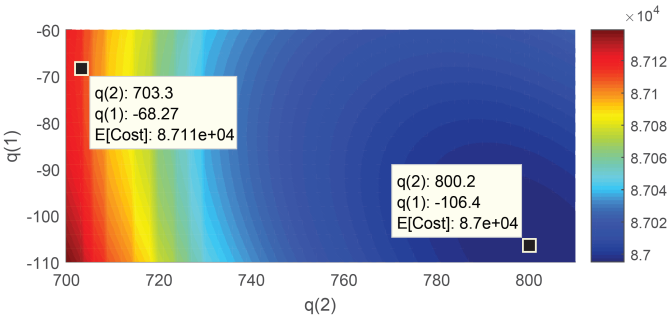
	Interface flows (MW)	LMP ($)	$\mathbb{E}$[Cost] ($)
CE	$q(1) = -68.27$ $q(2) = 703.32$	$\bar{\pi}_1 = 32.95$ $\bar{\pi}_2 = 32.94$ $\bar{\pi}_3 = 31.33$	87114.2
SIBIS	$q(1) = -106.42$ $q(2) = 800.16$	$\bar{\pi}_1 = 32.44$ $\bar{\pi}_2 = 32.44$ $\bar{\pi}_3 = 32.44$	86995.1

ing area was $32.94. Note that there was adequate transmission capacity, there was still extra transmission capacity and overall cost could have been reduced. This phenomenon is called interface under-utilization, which means transferring more power across the interface can further reduce the overall system cost. The economic benefit from SIBIS can be observed in the expected overall system cost reduction in the SIBIS schedule. By increasing the interface flow from area 3 to area 2 to 800.16 MW, the expected supply and demand prices converged to $32.44. Because there was no interface congestion, the expected prices in all three areas converged implying that the SIBIS schedule was efficient.

The convergence behavior of SIBIS is presented in Figure 4.13a where the expected overall cost was reasonably close to the optimum after the first two iterations. To demonstrate the optimality of the SIBIS schedule, we computed the expected costs in its neighborhood shown in Figure 4.13b where the SIBIS schedule is indicated by the cursor at the right bottom and the CE schedule at the left top. Note that the SIBIS schedule is located at the darkest point in this expected cost map which verifies the SIBIS converges to the globally optimal solution.



(a) Convergence



(b) Optimality

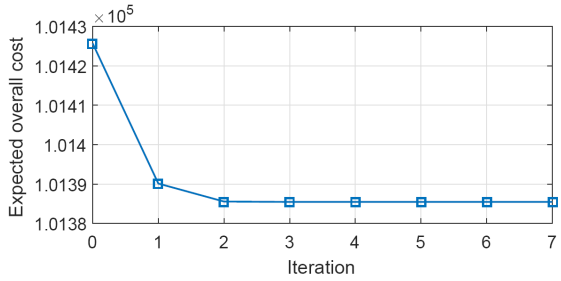
Figure 4.13: Convergence behavior and optimality of the synchronized scheduling algorithm in interface under-utilization the scenario.

Table 4.6: Performance comparison in the counter-intuitive flow scenario

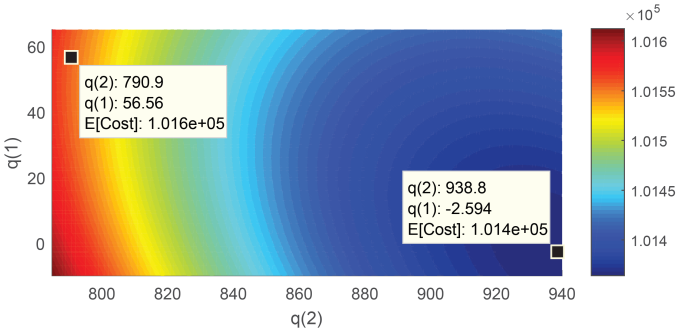
	Interface flows	LMP ($)	$\mathbb{E}$[Cost] ($)
CE	$q(1) = 56.56$ $q(2) = 790.86$	$\bar{\pi}_1 = 35.07$ $\bar{\pi}_2 = 34.80$ $\bar{\pi}_3 = 32.32$	101555.2
SIBIS	$q(1) = -2.59$ $q(2) = 938.78$	$\bar{\pi}_1 = 34.05$ $\bar{\pi}_2 = 34.05$ $\bar{\pi}_3 = 34.05$	101364.2

Simulation 2: Direction of Tie-line Flow

In this simulation, we examined the direction of power flow on tie lines. In this simulation, the same setting in Simulation 1 was used except that all loads in area 2 were increased by 20%.



(a) Convergence



(b) Optimality

Figure 4.14: Convergence behavior and optimality of the synchronized scheduling algorithm in the counter-intuitive flow scenario.

From the results presented in Table 4.6, the interface flows scheduled by the CE method and SIBIS were different in both directions and volumes. Note that the CE method scheduled 56.56 MW from area 1 to area 2: power flowed from higher- to lower-priced areas. This is economically counter-intuitive. In contrast, the SIBIS schedule resulted in price convergence between proxy buses, thus fundamentally eliminated counter-intuitive flows.

The convergence behavior was similar to that in Simulation 1 as shown in Figure 4.14a. The optimality was also verified.

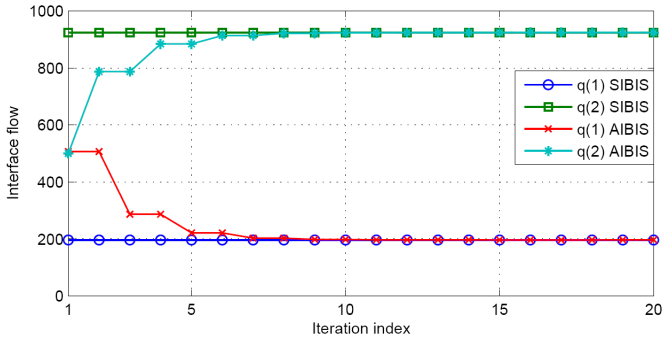
Asynchronous Interchange Scheduling

We have established theoretical results for AIBIS in Theorem 8 under the independent and stationary assumption for the net load process. In reality, however, the operating condition of the power grid is constantly changing. So we first verify the optimality of AIBIS in the i.i.d. case and compare the performance of SIBIS and AIBIS in a more general setting.

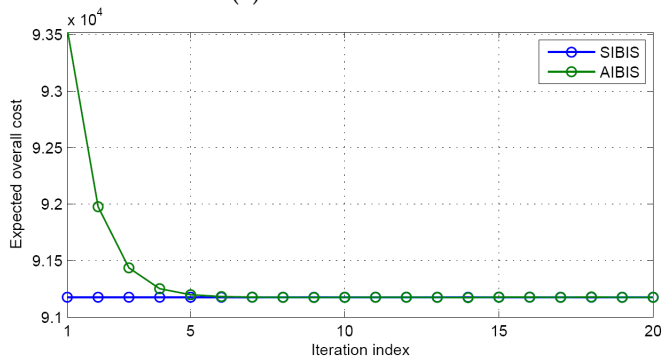
To verify the optimality of AIBIS for the i.i.d. wind process, the generation $w^t(i)$ of wind farm i at time t was assumed to follow $\mathcal{N}(100, 10^2)$ for all i and t. The scheduling horizon T was set to 20 for AIBIS and the tolerance ϵ was set to 0.001 for SIBIS. The initial interface flows were set at $(500, 500)$ for both algorithms.

As shown in Figure 4.15, SIBIS achieves optimal interchange for at every time whereas AIBIS achieves the optimality over time. Specifically, the interface flows and the expected cost of SIBIS are optimal and remain constant over time. AIBIS generated a sequence that converged to the optimal interchange and expected cost at time 7. It should be noted that the convergence rate is highly dependent on the initial values. The convergence time of AIBIS is simply the adaptation of the cyclic coordinate descent method.

The behavior of the proposed AIBIS algorithm was then investigated using a time varying process of wind generation. Specifically, we varied the mean value of wind generation starting from 100 MW and ending at 140 MW with a constant increment 2 MW. Except for the varying mean, the rest of the setting



(a) Interface flows

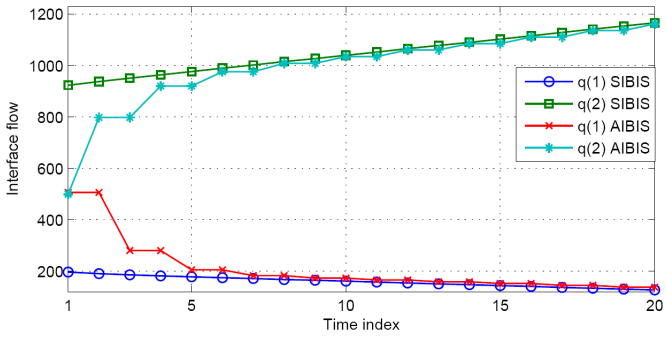


(b) Expected overall cost

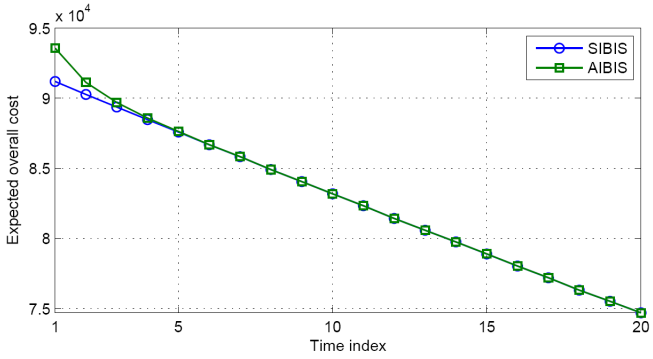
Figure 4.15: Performance comparison for SIBIS and AIBIS with i.i.d. wind generations.

remained the same.

From Figure 4.16a, the difference between SIBIS and AIBIS was clearly observed in the scheduled interface flows. Since there were only two interfaces, the interface flows scheduled by AIBIS were alternatively constant during a scheduling time slot. With time increasing, the scheduling difference decreased and so did the expected overall cost shown in Figure 4.16b.



(a) Interface flows



(b) Expected overall cost

Figure 4.16: Performance comparison for SIBIS and AIBIS with time-varying wind profiles.

CHAPTER 5

CONCLUSION

This thesis presents a new paradigm for power system operation under uncertainty. Two main components in this stochastic viewpoint are (1) the characterization of system uncertainty and (2) the model-specific solution method incorporating the probabilistic description of uncertainties. A parametric framework is proposed for a broad set of problems of optimization under uncertainty. The key idea of the proposed parametric framework is to have the result of the optimization pre-computed and stored for each parameter in the form as an algebraic function which can be easily evaluated. In other words, the optimal solution of the parameter is explicitly determined and not just implicitly as the result of an optimization problem. In this way, the operation decisions can be computed through the optimal mappings without repetitively solving the original optimization problems when the random variables are sampled or realized.

To illustrate the stochastic paradigm and the proposed parametric framework, we explored two specific problems in the real-time operations. First, we consider the short-term *probabilistic forecasting* of real-time LMP and congestion for system operators. Based on a multiparametric programming formulation of DCOPF, we have developed an approach that exploits the parametric structure of DCOPF solutions to obtain conditional distributions of future LMP and congestion. In this approach, the problem of collecting statistics on the space of *continuous probability distributions* of random parameters is reduced to that on the space of *finite discrete probability distributions* on a set of critical regions.

The approach presented in this thesis represents a first step toward online forecasting in large power systems with significant stochastic components. For system operators, the proposed online forecasting technique provides a new tool for managing operation risks and solving stochastic optimization problems. For market participants, on the other hand, congestion and LMP forecasts by the operator provide actionable signal for managing flexible resources and demand side management.

The second problem we consider is the stochastic interchange scheduling in multi-area systems. A stochastic interchange scheduling technique that incorporates load and renewable generation uncertainties is proposed based on proxy bus system. Using the forecast of the expected LMP at proxy buses, the proposed approach obtains the optimal interchange from an expected economic surplus maximization for the single interface flow. The essence of this technique is to convert a two-stage stochastic program into a deterministic optimization problem with a one-dimensional decision. Such a transformation makes the solution not only tractable but computationally efficient in the real-time market. In addition, the proposed technique does not require any iteration between operators during the scheduling process. A one-time information exchange is sufficient for the optimal scheduling.

For the generalization to multi-area multi-interface systems, a new technique is proposed based on the idea of coordinate descent method. The interchange vector is iteratively determined, one at a time, to minimize the expected overall system cost. The technique is specialized to both synchronous and asynchronous interchange scheduling. The optimality and convergence are guaran-

teed under nominal assumptions.

In summary, this thesis presents a new paradigm for operations under uncertainty in the modern electric grid. An efficient stochastic framework is proposed to investigate the effect of uncertainty on the optimal operation based on the multiparametric programming theory. This parametric framework provides a computational efficient methodology for characterizing system uncertainty and incorporating the probabilistic distribution into operations.

APPENDIX A

PROOFS

A.1 Proofs for Chapter 2

A.1.1 Proof of Theorem 1

For any parameter θ, if θ is in the same critical region as θ_0, they have the same optimal partition, which means that

$$A_{\mathcal{J}_0}x^*(\theta) - b_{\mathcal{J}_0} - E_{\mathcal{J}_0}\theta = \mathbf{0}, \tag{A.1}$$

$$A_{\mathcal{J}_0^c}x^*(\theta) - b_{\mathcal{J}_0^c} - E_{\mathcal{J}_0^c}\theta < \mathbf{0}. \tag{A.2}$$

Because MPLP (2.2) is assumed to be neither primal nor dual degenerate, $A_{\mathcal{J}_0}$ has full rank, and

$$x^*(\theta) = A_{\mathcal{J}_0}^{-1}(b_{\mathcal{J}_0} + E_{\mathcal{J}_0}\theta). \tag{A.3}$$

Substituting $x^*(\theta)$ into (A.2), we have $\theta \in \Theta_{\mathcal{J}_0}$; substituting $x^*(\theta)$ into the objective function $z^*(\theta) = z(x^*(\theta)$, we have (2.6).

Since $x^*(\theta)$ and $y^*(\theta) = y^*(\theta_0)$ satisfy the optimal conditions for (2.2), the optimal dual solution is a constant vector $y^*(\theta_0)$ for all $\theta \in \Theta_{\mathcal{J}_0}$.

A.1.2 Proof of Theorem 2

The piecewise affine property of optimizer and value function follows immediately from Theorem 1 and from the enumeration of all possible combinations of active constraint sets. Convexity of $z^*(\cdot)$ and continuity of $x^*(\cdot)$ follows from standard results on multiparametric programs.

A.1.3 Proof of Theorem 3

The first order optimality conditions for MPQP (2.7) are given by

$$\begin{aligned} Hx^*(\theta) + A^\intercal y^*(\theta) &= \mathbf{0}, && \text{(A.4)} \\ Ax^*(\theta) - b - E\theta &\leq \mathbf{0}, && \text{(A.5)} \\ y_i^*(\theta)(A_i x^*(\theta) - b_i - E_i\theta) &= 0, \ \forall i \in \mathcal{J}, && \text{(A.6)} \\ y^*(\theta) &\geq \mathbf{0}. && \text{(A.7)} \end{aligned}$$

From (A.4),

$$x^*(\theta) = -H^{-1}A^\intercal y^*(\theta). \qquad \text{(A.8)}$$

Substituting the result into (A.6), we have

$$\begin{aligned} A_i H^{-1} A_i^\intercal y_i^*(\theta) + b_i + E_i\theta &= 0, \ \forall i \in \mathcal{I}_0, && \text{(A.9)} \\ y_i^*(\theta) &= 0, \ \forall i \in \mathcal{I}_0^c. && \text{(A.10)} \end{aligned}$$

By the non-degeneracy assumption, the rows of $A_{\mathcal{I}_0}$ are linearly independent. This implies that $A_{\mathcal{I}_0} H^{-1} A_{\mathcal{I}_0}^\intercal$ is a square full rank matrix. Therefore, from (A.9), we solve

$$y_{\mathcal{I}_0}^*(\theta) = -(A_{\mathcal{I}_0} H^{-1} A_{\mathcal{I}_0}^\intercal)^{-1}(b_{\mathcal{I}_0} + E_{\mathcal{I}_0}\theta) \qquad \text{(A.11)}$$

and substitute $y^*_{\mathcal{J}_0}(\theta)$ and $y_{\mathcal{J}_0^c}(\theta)$ into (A.8) to obtain

$$x^*(\theta) = H^{-1}A_{\mathcal{J}_0}^\intercal(A_{\mathcal{J}_0}H^{-1}A_{\mathcal{J}_0}^\intercal)^{-1}(b_{\mathcal{J}_0} + E_{\mathcal{J}_0}\theta). \tag{A.12}$$

Substituting $x^*(\theta)$ from (2.9) in the primal feasibility conditions (A.5) gives $\mathcal{P}_p$ and substituting $x^*(\theta)$ from (2.9) in the dual feasibility condition (A.7) gives $\mathcal{P}_d$. We therefore have $\theta \in \Theta_{\mathcal{J}_0}$.

A.1.4 Proof of Theorem 4

The piecewise affine property of optimizer and the piecewise quadratic property of value function follow immediately from Theorem 3 and from the enumeration of all possible combinations of active constraint sets. Convexity and continuous differentiability of $z^*(\cdot)$ and continuity of $x^*(\cdot)$ follows from standard results on multiparametric programs and non-degeneracy assumption.

A.2 Proofs for Chapter 3

A.2.1 Proof of Theorem 5

We first derive the exact piece-wise affine expression of $\pi(\theta)$. Assume the quadratic generation cost function is given by $g^\intercal Hg + c^\intercal g$ where H is a diagonal matrix with all values positive, the parametric DCOPF (3.1) can be written

in the succinct form of (2.2) as follows

$$\min_{g} g^\intercal H g + c^\intercal g \text{ subject to } Ag \le b + E\theta \quad (\gamma) \tag{A.13}$$

where $\gamma = (\lambda, -\lambda, \mu^+, \mu^-, \nu^+, \nu^-)$ is the vector of all Lagrangian multipliers[1], and the coefficient matrices A, E, and vector b are given by

$$A = \begin{bmatrix} \mathbf{1}^\intercal \\ -\mathbf{1}^\intercal \\ S \\ -S \\ I_n \\ -I_n \end{bmatrix}, E = \begin{bmatrix} \mathbf{1}^\intercal & \mathbf{1}^\intercal \\ -\mathbf{1}^\intercal & -\mathbf{1}^\intercal \\ S & S \\ -S & -S \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ L^+ \\ -L^- \\ G^+ \\ -G^- \end{bmatrix}.$$

For a given critical region Θ_i, let $(\mathcal{J}_i, \mathcal{J}_i^c)$ denote the optimal partition. By (A.11), the Lagrangian multipliers $\gamma_{\mathcal{J}_i}$ associated with $\mathcal{J}_i$ is given by

$$\gamma_{\mathcal{J}_i}(\theta) = -(A_{\mathcal{J}_i} H^{-1} A_{\mathcal{J}_i}^\intercal)^{-1} (E_{\mathcal{J}_i}\theta + b_{\mathcal{J}_i}) \tag{A.14}$$

where $A_{\mathcal{J}_i}$, $E_{\mathcal{J}_i}$ and $b_{\mathcal{J}_i}$ are the submatrices of A, E and subvector of b corresponding to $\mathcal{J}_i$.

Let W_i be the submatrix of $-(A_{\mathcal{J}_i} H^{-1} A_{\mathcal{J}_i}^\intercal)^{-1}$ consisting of rows associated with the power balance constraint and the active transmission constraints. Then we have

$$\begin{bmatrix} \lambda(\theta) \\ \bar{\mu}(\theta) \end{bmatrix} = W_i (E_{\mathcal{J}_i}\theta + b_{\mathcal{J}_i}) \tag{A.15}$$

where $\bar{\mu}(\theta) = (-\bar{\mu}_+(\theta), \bar{\mu}_-(\theta))$ and $\bar{\mu}_+(\theta)$ and $\bar{\mu}_-(\theta)$ are Lagrangian multipliers associated with active transmission constraints.

[1]The time indices are removed for notational convenience.

The LMP function $\pi_i(\theta)$ for $\theta \in \Theta_i$ is given by

$$\pi_i(\theta) = \mathbf{1}\lambda(\theta) - S^\intercal \mu^+(\theta) + S^\intercal \mu^-(\theta) \tag{A.16}$$

$$= \begin{bmatrix} \mathbf{1} & \bar{S}_i^\intercal \end{bmatrix} \begin{bmatrix} \lambda(\theta) \\ \bar{\mu}(\theta) \end{bmatrix} \tag{A.17}$$

$$= \begin{bmatrix} \mathbf{1} & \bar{S}_i^\intercal \end{bmatrix} W_i \left(E_{\mathcal{I}_i} \theta + b_{\mathcal{I}_i} \right) \tag{A.18}$$

$$= U_i \theta + v_i \tag{A.19}$$

where $U_i = \begin{bmatrix} \mathbf{1} & \bar{S}_i^\intercal \end{bmatrix} W_i E_{\mathcal{I}_i}$ and $v_i = \begin{bmatrix} \mathbf{1} & \bar{S}_i^\intercal \end{bmatrix} W_i E_{\mathcal{I}_i} b_{\mathcal{I}_i}$.

Because of the non-degeneracy assumption of (3.1) (or equivalently (2.7)), $\lambda(\theta)$ and $\bar{\mu}(\theta)$ are unique for θ. Since matrix $\begin{bmatrix} \mathbf{1} & \bar{S}_i^\intercal \end{bmatrix}$ has full column rank, $\pi_i(\theta)$ is one-to-one correspondence from Θ_i to Π_i.

As an immediate result of the following lemma, the PDF of the LMP π is given by (3.19).

Lemma 1. *Let $X \in \mathcal{X} \subseteq \mathbb{R}^m$ be a continuous random vector with PDF $f_X(x)$ and $\{\mathcal{X}_k\}_{k=1}^K$ be a collection of sets that partition $\mathcal{X}$. Define a piecewise transformation $L : \mathbb{R}^m \to \mathbb{R}^n$*

$$L(x) = L_k(x), x \in \mathcal{X}_k$$

where $L_k : \mathcal{X}_k \to \mathbb{R}^n$ is assumed to be injective.

The PDF of $Y = L(X)$ is given by

$$f_Y(y) = \sum_{k=1}^{K} \mathbb{I}_{\mathcal{Y}_k}(y) f_k(y) \tag{A.20}$$

where $\mathcal{Y}_k$ is the image of L_k on $\mathcal{X}_i$ and $f_k(y)$ the PDF of random variable $Z = L_k(X)$.

A.2.2 Proof of Lemma 1

Let $\mathcal{Y}_\epsilon(y) = \{y' | \parallel y' - y \parallel < \epsilon\}$ for all y. Since $\{\mathcal{X}_k\}_{k=1}^K$ partition $\mathcal{X}$, and L_k is injective for all k, for sufficiently small $\epsilon > 0$,

$$\mathbb{P}\left[Y \in \mathcal{Y}_\epsilon(y)\right] = \sum_{k=1}^{K} \mathbb{I}_{\mathcal{X}_k}(L_k^{-1}(y))\mathbb{P}\left[X \in L_k^{-1}(\mathcal{Y}_\epsilon(y))\right] \tag{A.21}$$

$$= \sum_{k=1}^{K} \mathbb{I}_{\mathcal{Y}_k}(y)\mathbb{P}\left[Y = L_k(X) \in \mathcal{Y}_\epsilon(y)\right] \tag{A.22}$$

$$= \sum_{k=1}^{K} \int_{\mathcal{Y}_\epsilon(y)} \mathbb{I}_{\mathcal{Y}_k}(y) f_i(z)dz \tag{A.23}$$

$$\tag{A.24}$$

Therefore, for the image $\mathcal{Y}$ of L on $\mathcal{X}$,

$$\int_{\mathcal{Y}} f_Y(y)dy = \int_{\mathcal{Y}} \sum_{k=1}^{K} \mathbb{I}_{\mathcal{Y}_k}(y) f_i(y)dy. \tag{A.25}$$

A.3 Proofs for Chapter 4

A.3.1 Proof of Theorem 6

We first show the differentiability of the objective functions of (4.6) and (4.8). This follows immediately from the multiparametric quadratic programming results summarized in Theorem 4.

By Theorem 4, the objective function of (4.6), denoted by $J_6(q)$, is continu-

ously differentiable with derivative

$$\begin{aligned} J_6'(q) &= \mathbb{E}_{d_1,d_2}\left[\frac{\partial}{\partial q}c_1(g_1^*(q,d_1)) + \frac{\partial}{\partial q}c_2(g_2^*(q,d_2))\right] \\ &= \bar{\pi}_1(q) - \bar{\pi}_2(q) \end{aligned}$$

where the second equality holds by the Envelope Theorem. Since $\pi_1(q, d_1)$ and $\pi_2(q, d_2)$ are continuous functions, the objective function of (4.8), denoted by $J_8(q)$, is differentiable with derivative

$$J_8'(q) = \bar{\pi}_2(q) - \bar{\pi}_1(q).$$

We then derive the connection between the optimal solutions to (4.6) and (4.8) using the first order conditions. The optimal solution q^* to (4.6) and associated Lagrangian multiplier μ_q^* (for the interface constraint) satisfy the following first order condition for (4.6):

$$\bar{\pi}_1(q^*) - \bar{\pi}_2(q^*) + \mu_q^* = 0. \tag{A.26}$$

Similarly, the optimal solution $q^\sharp$ to (4.8) and associated Lagrangian multiplier $\mu_q^\sharp$ (for the interface constraint) satisfy the following first order condition for (4.8):

$$\bar{\pi}_1(q^\sharp) - \bar{\pi}_2(q^\sharp) + \mu_q^\sharp = 0, \tag{A.27}$$

which is exactly the same as (A.26).

Finally, we show $q^* = q^\sharp$. To prove this, we need the monotonicity of price function $\bar{\pi}_i(q)$ which is summarized in the following lemma.

Lemma 2. *If problem (4.3) and (4.4) are neither primal nor dual degenerate for all d_1, d_2, and $q \le Q$, then $\bar{\pi}_1(q)$ is monotonically increasing and $\bar{\pi}_2(q)$ monotonically decreasing, where the direction of q is defined from region 1 to region 2.*

Proof. See Appendix A.3.2. □

Below we show that in each of the following cases, either the case itself is impossible or $q^* = q^\sharp$.

(1) $q^* = q^\sharp = Q$. The statement is trivially true.

(2) $q^* < Q$ and $q^\sharp < Q$. In this case, the interface constraint is not binding in either problem, so we have $\mu_q^* = \mu_q^\sharp = 0$, which implies that $\bar{\pi}_1(q^*) = \bar{\pi}_2(q^*)$ and $\bar{\pi}_1(q^\sharp) = \bar{\pi}_2(q^\sharp)$. By the monotonicity of $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$, $q^* = q^\sharp = q_{\text{int}}$.

(3) $q^* < q^\sharp = Q$. In this case, $\mu^* = 0$ and $\bar{\pi}_1(q^*) = \bar{\pi}_2(q^*)$. Because of the monotonicity of $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$, $J_8(q)$ is decreasing in $q \in (q_{\text{int}}, Q]$. Since $q^\sharp > q^* = q_{\text{int}}$, we have $J_8(q^*) > J_8(q^\sharp)$. This contradicts with the optimality of $q^\sharp$. Therefore, this case is impossible.

(4) $q^\sharp < q^* = Q$. This case is also impossible. The proof follows the logic of that in Case (3).

To sum up, problems (4.6) and (4.8) have the same optimal solution $q^* = q^\sharp = \min(q_{\text{int}}, Q)$.

A.3.2 Proof of Lemma 2

Let the Lagrangian functions for (4.3) and (4.4) be $\mathcal{L}_3$ and $\mathcal{L}_4$ respectively. By Theorem 4, $c_i^*(q, d_i) \triangleq c_i(g_i^*(q, d_i))$, $i \in \{1, 2\}$, is continuously differentiable. By the

Envelope Theorem,

$$\frac{\partial c_1(g_1^*(q,d_1))}{\partial q} = \frac{\partial \mathcal{L}_3}{\partial q} = \pi_1(q,d_1)$$
$$\frac{\partial c_2(g_2^*(q,d_2))}{\partial q} = \frac{\partial \mathcal{L}_4}{\partial q} = -\pi_2(q,d_2)$$

where $\pi_1(q,d_1)$ and $\pi_2(q,d_2)$ are continuous functions.

Since $c_i^*(q,d_i)$ is also convex, its second order partial derivative of $c_i^*(q,d_i)$ with respect to q is positive. Therefore, for fixed d_1 and d_2, $\pi_1(q,d_1)$ is monotonically increasing and $\pi_2(q,d_2)$ monotonically decreasing for $q \leq Q$. Note that monotonicity is preserved under expectation, so $\bar{\pi}_1(q)$ and $\bar{\pi}_2(q)$ are also monotonic in q.

A.3.3 Proof of Theorem 7

We first give the properties, implied by Theorem 4, of the objective function $\bar{C}(q)$ in (4.10) and establish the convergence of the algorithm. Let $q^{(k_i)} = \left(q^{(k)}(i), q^{(k)}(-i)\right)$ where $q^{(k)}(-i)$ is defined in (4.15). Using the updating rule (4.14), we have

$$\bar{C}(q^{(k)}) \leq \bar{C}(q^{(k_{I-1})}) \leq \cdots \leq \bar{C}(q^{(k_1)}) \leq \bar{C}(q^{(k-1)}), \forall k. \tag{A.28}$$

Let q^* be a limit point of the sequence $\{q^{(k)}\}_{k=1}^{\infty}$. Note that $q^* \in \mathcal{Q} = \{q | q \leq Q\}$ because $\mathcal{Q}$ is closed. The monotonicity (A.28) implies that the sequence $\{\bar{C}(q^{(k)})\}_{k=1}^{\infty}$ converges to $\bar{C}(q^*)$.

Let $\{q^{(k)}\}_{k\in\mathcal{K}}$, where $\mathcal{K}$ is an index set, be a subsequence of $\{q^{(k)}\}_{k=1}^{\infty}$ that converges to q^*. From the updating rule (4.14) and the monotonic property (A.28),

for any interface i, we have

$$\bar{C}\left(q^{(k)}\right) \leq \bar{C}\left(q^{(k_i)}\right) \leq \bar{C}\left(q(i), q^{(k)}(-i)\right), \forall q(i) \leq Q(i).$$

Since $\bar{C}(q)$ is continuous, implied by Theorem 4, taking the limit as $k \in \mathcal{K}$ tends to infinity on both sides, we have

$$\bar{C}(q^*) \leq \bar{C}\left(q(i), q^*(-i)\right), \forall q(i) \leq Q(i)$$

which means $q^*(i)$ is an optimal solution of the following optimization

$$\min_{q(i) \leq Q(i)} \bar{C}\left(q(i), q^*(-i)\right). \tag{A.29}$$

Therefore, $q^*(i)$ satisfies the Karush-Kuhn-Tucker (KKT) conditions for (A.29), *i.e.*

$$\nabla_i \bar{C}(q^*(i), q^*(-i)) + \lambda(i)(q^*(i) - Q(i)) = 0 \tag{A.30}$$

$$q^*(i) \leq Q(i) \tag{A.31}$$

$$\lambda(i) \geq 0 \tag{A.32}$$

where $\nabla_i \bar{C}(q)$ is the partial derivative with respective to $q(i)$ and $\lambda(i)$ the associated Lagrangian multiplier.

Note that conditions (A.30-A.32) hold for all i at q^*, *i.e.*,

$$\nabla \bar{C}(q^*) + \lambda(q^* - Q) = 0 \tag{A.33}$$

$$q^* \leq Q \tag{A.34}$$

$$\lambda \geq 0. \tag{A.35}$$

Since conditions (A.33-A.35) are the KKT conditions for (4.10), and $\bar{C}(q)$ is convex by Theorem 4, q^* is optimal to (4.10).

A.3.4 Proof of Theorem 8

Since $d_n^t \overset{i.i.d}{\sim} \mathcal{F}_n$ for all area n, $q^t(i)$ defined in (4.16) can be obtained by

$$q^t(i) = \underset{q(i)\leq Q(i)}{\arg\min} \bar{C}\left(q(i), q^{t-1}(-i)\right) \tag{A.36}$$

which implies

$$\bar{C}(q^t) \leq \bar{C}(q^{t-1}), \forall t.$$

Let $\tilde{q}$ be a limit point of the sequence $\{q^t\}_{t=1}^{\infty}$. The monotonicity of $\bar{C}(q^t)$ implies that the sequence $\{\bar{C}(q^t)\}_{t=1}^{\infty}$ converges to $\bar{C}(\tilde{q})$.

Let $\{q^t\}_{t\in\mathcal{T}}$, where $\mathcal{T}$ is an index set, be a subsequence of $\{q^t\}_{t=1}^{\infty}$ that converges to $\tilde{q}$. From the updating rule (A.36) and the monotonicity of $\bar{C}(q^t)$, we have

$$\bar{C}\left(q^t\right) \leq \bar{C}\left(q(i), q^{t-1}(-i)\right), \forall t, \forall i, \forall q(i) \leq Q(i).$$

Since $\bar{C}(q)$ is continuous, implied by Theorem 4, taking the limit as $t \in \mathcal{T}$ tends to infinity on both sides, we have

$$\bar{C}(\tilde{q}) \leq \bar{C}\left(q(i), \tilde{q}(-i)\right), \forall q(i) \leq Q(i)$$

which means $\tilde{q}(i)$ is an optimal solution of the following optimization

$$\min_{q(i)\leq Q(i)} \bar{C}\left(q(i), \tilde{q}(-i)\right). \tag{A.37}$$

Therefore, $\tilde{q}(i)$ satisfies the KKT conditions (A.30-A.32) for A.37 at $\tilde{q}$. Since (A.30-A.32) hold for all i at $\tilde{q}$, $\tilde{q}$ satisfies the KKT conditions (A.33-A.35). By the convexity of $\bar{C}(q)$, the KKT conditions are sufficient and necessary for optimality. Therefore, $\tilde{q}$ is optimal to (4.10).

BIBLIOGRAPHY

[1] TSC, Intraday Congestion Forecast (IDCF), http://www.amprion.net/pressemitteilung-126.

[2] PSE S.A., http://www.psi.de/en/psi-pressevents/releases-archive/article/article/psi-receives-contract-for-implementing-intraday-congestion-forecast-for-polish-transmission-system-o/.

[3] PJM, http://www.pjm.com/.

[4] NYISO,http://www.nyiso.com/public/index.jsp.

[5] ISO New England, http://www.iso-ne.com/.

[6] Governor cuomo announces establishment of clean energy standard that mandates 50 percent renewables by 2030. https://www.governor.ny.gov/news/governor-cuomo-announces-establishment-clean-energy-standard-mandates-50-percent-renewables. Accessed: 2016-11-17.

[7] Ieee 118 bus system. motor.ece.iit.edu/data/JEAS_IEEE118.doc.

[8] Ilan Adler and Renato DC Monteiro. A geometric view of parametric linear programming. *Algorithmica*, 8(1-6):161–176, 1992.

[9] AESO. Pool price forecast calculation methodology. http://www.aeso.ca/downloads/Price_Forecast_Calculation_March_8_2011.pdf.

[10] Ali Ahmadi-Khatir, Antonio J Conejo, and Rachid Cherkaoui. Multi-area energy and reserve dispatch under wind uncertainty and equipment failures. *IEEE Transactions on Power Systems*, 28(4):4373–4383, 2013.

[11] C Auxenfans, R Baumann, T Lapetanovic, P D Leener, R Paprocki, A Wirth, and D Klaar. European transmission system operators' cooperation for enhanced system security, integration of renewables and market support. In *Proc. of the Council on Large Electric Systems (CIGRE)*, 2014.

[12] Ross Baldick and Dhiman Chatterjee. Coordinated dispatch of regional transmission organizations: Theory and example. *Computers & Operations Research*, 41:319–332, 2014.

[13] Bernd Bank, Jürgen Guddat, Diethard Klatte, and Bernd Kummer. *Nonlinear parametric optimization.* Birkhauser Verlag, Kinokuniya Company Ltd., 1983.

[14] Alberto Bemporad, Manfred Morari, Vivek Dua, and Efstratios N Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3–20, 2002.

[15] F. Borrelli, A. Bemporad, and M. Morari. Geometric algorithm for multiparametric linear programming. *Journal of Optimization Theory and Applications*, 118(3):515–540, 2003.

[16] F Borrelli, A Bemporad, and M Morari. Predictive control for linear and hybrid systems. `http://www.mpc.berkeley.edu/mpc-course-material`, 2015.

[17] GLENN W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

[18] Michael D Cadwalader, Scott M Harvey, Susan L Pope, and William W Hogan. Market coordination of transmission loading relief across multiple regions. *Cambridge, MA: Center for Business and Government, Harvard University*, 1998.

[19] Jie Chen, James S Thorp, and Timothy D Mount. Coordinated interchange scheduling and opportunity cost payment: a market proposal to seams issues. In *Proc. of the 37th Annual Hawaii International Conference on System Sciences*, 2004.

[20] Antonio J Conejo and Jose A Aguado. Multi-area coordinated decentralized dc optimal power flow. *IEEE Transactions on Power Systems*, 13(4):1272–1278, 1998.

[21] Meysam Doostizadeh, Farrokh Aminifar, Hamid Lesani, and Hassan Ghasemi. Multi-area market clearing in wind-integrated interconnected power systems: A fast parallel decentralized method. *Energy Conversion and Management*, 113:131–142, 2016.

[22] ERCOT. Wholesale price forecast tool.

[23] T. Gal and J. Nedoma. Multiparametric linear programming. *Management Science*, 18:406–442, 1972.

[24] Tomas Gal. *Postoptimal analyses, parametric programming and related topics*. Walter de Gruyter, 1995.

[25] Tomas Gal and Harvey J Greenberg. *Advances in sensitivity analysis and parametric programming*, volume 6. Springer Science & Business Media, 2012.

[26] Saul Gass and Thomas Saaty. The computational algorithm for the parametric objective function. *Naval research logistics quarterly*, 2(1-2):39–45, 1955.

[27] Global Energy Forecasting Competition, 2014.

[28] S Harvey. Proxy buses, seams and markets [draft], 2003. `http://www.hks.harvard.edu/hepg/Papers/Harvey_Proxy.Buses.Seams.Markets_5-23-03.pdf`.

[29] Kory W Hedman, Shmuel S Oren, and Richard P O'Neill. A review of transmission switching and network topology optimization. In *Proc. of IEEE Power and Energy Society General Meeting,*, pages 1–7, 2011.

[30] M. Herceg, M. Kvasnica, C.N. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510, Zürich, Switzerland, July 17-19 2013. `http://people.ee.ethz.ch/~mpt/3/`.

[31] Clifford Hildreth. A quadratic programming procedure. *Naval research logistics quarterly*, 4(1):79–85, 1957.

[32] ISO New England. Real-time actual scheduled interchange.

[33] Y. Ji, J. Kim, R. J. Thomas, and L. Tong. Forecasting real-time locational marginal price: A state space approach. In *Proc. of the 47th Asilomar Conference on Signals, Systems, and Computers*, pages 379–383, 2013.

[34] Yuting Ji and Lang Tong. Multi-area interchange scheduling under uncertainty. *IEEE Transactions on Power Systems (submitted)*.

[35] Balho H Kim and Ross Baldick. Coarse-grained distributed optimal power flow. *IEEE Transactions on Power Systems*, 12(2):932–939, 1997.

[36] Kenneth Kreutz-Delgado, Joseph F Murray, Bhaskar D Rao, Kjersti Engan, Tai Sing Lee, and Terrence J Sejnowski. Dictionary learning algorithms for sparse representation. *Neural computation*, 15(2):349–396, 2003.

[37] Zhigang Li, Wenchuan Wu, Mohammad Shahidehpour, and Boming Zhang. Adaptive robust tie-line scheduling considering wind power uncertainty for interconnected power systems. *IEEE Transactions on Power Systems*, 31(4):2701–2713, 2016.

[38] Zhi-Quan Luo and Paul Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72(1):7–35, 1992.

[39] T Mai, R Wiser, D Sandor, G Brinkman, G Heath, P Denholm, DJ Hostick, N Darghouth, A Schlosser, and K Strzepek. Exploration of high-penetration renewable electricity futures. vol. 1 of renewable electricity futures study. *National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/TP-6A20-52409-1*, 2012.

[40] Liang Min, Stephen T. Lee, Pei Zhang, Virgil Rose, and James Cole. Short-term probabilistic transmission congestion forecasting. In *Proc. of the 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 764–770, 2008.

[41] New York ISO. External rto cts price.

[42] Andrew L. Ott. Experience with pjm market operation, system design, and implementation. *IEEE Transactions on Power Systems*, 18(2):528–534, 2003.

[43] Efstratios N Pistikopoulos, Luis Dominguez, Christos Panos, Konstantinos Kouramas, and Altannar Chinchuluun. Theoretical and algorithmic advances in multi-parametric programming and control. *Computational Management Science*, 9(2):183–203, 2012.

[44] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[45] Paul Tseng. Convergence of a block coordinate descent method for non-

differentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001.

[46] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014.

[47] Matthew White and Robert Pike. Inter-regional interchange scheduling analysis and options. *ISO New England and New York ISO, ISO White Paper*, 2011.

[48] Feng Zhao, Eugene Litvinov, and Tongxin Zheng. A marginal equivalent decomposition method and its application to multi-area optimal power flow problems. *IEEE Transactions on Power Systems*, 1(29):53–61, 2014.

[49] Tongxin Zheng and Eugene Litvinov. Ex post pricing in the co-optimized energy and reserve market. *IEEE Transactions on Power Systems*, 21(4):1528–1538, 2006.

[50] Qun Zhou, Leigh Tesfatsion, and Chen-Ching Liu. Short-term congestion forecasting in wholesale power markets. *IEEE Transactions on Power Systems*, 26(4):2185–2196, 2011.

[51] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, 26(1):12–19, 2011.